THE UNIVERSITY *of* EDINBURGH

Use of 2-Aminopurine Fluorescence as a Probe of DNA and Computational Studies of a New Class of Base Analogues

Xiaohua Wu

Degree of Doctor of Philosophy

The University of Edinburgh

2012

Abstract

The steady-state and time-resolved fluorescence of 2-aminopurine (2AP) have been used to monitor base dynamics and base stacking interactions in DNA single strands and dinucleotides, and to investigate the interactions between DNA and a polymerase, Pfu-Pol. A new class of base analogues has also been investigated using a combination of experiment and quantum chemical computation.

In recent years, 2AP has been widely used as a fluorescent probe to study conformational changes and inter-bases interactions in duplex DNA, but the conformational behaviour of DNA in single strands has been far less investigated. In the present work, six 2AP–labelled single strands have been studied by steady-state and time-resolved fluorescence measurements. Single strands were found to show similar conformational heterogeneity (manifested by 4-exponential fluorescence decays) to duplex DNA, but highly stacked conformations, in which 2AP is rapidly quenched by inter-base charge transfer, are less populated in single strands, whereas imperfectly stacked (weakly quenched) conformations are more highly populated. The effect of base pairing in constraining base mobility is evident.

To further investigate the influence of base stacking interaction on DNA conformation and the mechanism of inter-base quenching of 2AP, the time-resolved fluorescence of 2AP-containing dinucleotides was measured. The fluorescence decay of 2AP-containing dinucleotides in PBS buffer at room temperature is also multiexponential and the shortest lifetime varies with the identity of the natural base partner, in a manner consistent with quenching by inter-base electron transfer. When the dinucleotides are frozen to 77 K, the quenching of 2AP is almost eliminated, demonstrating the importance of thermal fluctuations of the bases in facilitating inter-base quenching at room temperature. In the frozen dinucleotides, an additional decay component with a lifetime significantly longer than unquenched 2AP is also observed, suggesting the formation of a new, delocalised, inter-base excited-state.

Archaeal family-B DNA polymerases bind tightly to uracil and stall replication when they encounter this base in template strands, four bases ahead of the primer-template junction. If the polymerase progresses further towards the uracil, the 3′-5′ proof reading exonuclease becomes stimulated, trimming the primer and re-setting uracil to the +4 position. Uracil sensing prevents copying of the deaminated base and the introduction of mutations into the genome. Time-resolved fluorescence of 2AP has been used to investigate the role played by unwinding of primer-templates in this mechanism. 2AP-labelled primer-templates (2AP positioned next to the terminal 3′ base of the primer strand), with a misincorpated uracil at the +2 position (U+2) or +4 position (U+4) from the replication fork in the complementary template strand, were investigated in complex with the polymerase Pfu-Pol. For the U+2 primer-template, the fluorescence decay parameters show clear evidence for a decrease in the amount of double-stranded DNA on polymerase binding, manifested by marked weakening of inter-base stacking and a large transfer of population from highly stacked to poorly stacked conformations. In contrast, for the U+4 primer-template only a small

perturbation to inter-base stacking is seen, together with the persistence of a high population of strongly stacked states.

A new class of base analogues with selenium replacing oxygen at the 4 position of thymine and the 6 position of guanine has been investigated experimentally and computationally. These base analogues are interesting because they show a large shift (>80 nm) in their absorption spectrum compared with the natural bases, taking their absorption into the visible region, with minimal change in molecular structure. The potential of two examples of these analogues, 4-selenium thymine-3'-phosphate and 6-selenium-2'-deoxyguanosine-3-phosphate as luminescent probes has been investigated. However, they prove to have very low emission quantum yields for both fluorescence and phosphorescence. The effect of selenium-substitution on the structural and photophysical properties of the bases has been studied by various *ab initio* computational methods. It has been found that replacement of oxygen by selenium does not affect the ground state structure but changes the structure of the first excited-state from buckled to nearly planar. The shift in the absorption spectrum on introduction of selenium is successfully predicted by the calculations; the red-shifted absorption band of selenium-substituted thymine is due to a new electronic transition that is not present in the natural base, whereas that of selenium-substituted guanine is from red-shifting of a guanine-like transition.

Declaration

I declare that the work presented in this thesis is my own unless otherwise stated by reference, and that the work has not been submitted for any other degree or professional qualification except as specified.

Signed

Date

Acknowledgements

My post-graduate studies would not have been finished without the help, support and encouragement from the following people.

Firstly, I would like to express my sincerest gratitude to my supervisor, Dr. Anita. C. Jones, for all her great knowledge, guidance, patience, and enormous support throughout my PhD. I would also like to thank my second supervisor Dr. David Dryden for his useful discussions and great knowledge in the area of molecular biology.

Secondly, I would like to thank Dr. Robert Neely for teaching me how to use the laser, carry out the experiments, and handle experimental data. Thanks also go to Dr. Patricia Richardson, who gave me much technical support in the computational calculations and guidance in writing the corresponding chapter in my thesis. I would like to thank the FLIM lab manager in the COSMIC, Dr. Jochen Arlt, who was always there when the laser and third harmonic generator did funny things. I would also like to thank the folks for their friendship, enjoyable times and exchange of knowledge during my years of study in the University of Edinburgh.

Thirdly, I would like to express my deeply gratitude to all my family for their endless love. In particular, my father Huanliang Wu and mother Lanying Wang for giving me the life, bringing me up and supporting all my dreams; and my husband Yulun Song for staying with me in UK, cooking delicious meals, huge encouragement, and brightening up my life.

Finally, acknowledgement goes to the EaStCHEM Research School and company MTEM for the funding of my PhD study.

Table of Contents

Chapter 1

Introduction

The aims of the work presented were to investigate the base stacking interactions and conformational changes in DNA single strands and dinucleotides, to investigate the mechanism of the action of a polymerase upon binding to DNA, and to explore a new class of base analogues.

Fluorescence spectroscopy, including steady-state and time-resolved fluorescence, is an excellent method to study the interactions and dynamics of biomolecules, because of the sensitivity and selectivity of this technique. The natural nucleic bases are not fluorescent because of their extremely low quantum yield. Therefore, in order to use fluorescence spectroscopy to study DNA, a fluorescent probe must be incorporated into the system. 2-aminopurine (2AP) has been widely used as fluorescent probe of DNA for nearly thirty years. It is an analogue of adenine and can be incorporated into DNA without significant structural perturbation to the duplex. Furthermore, the fluorescence properties of 2AP are sensitive to the local molecular environment, particular base-stacking interactions.

In this thesis, 2AP has been incorporated into DNA single strands, dinucleotides and the prime strand of DNA duplexes. The base stacking interactions and conformational changes in DNA single strands and dinucleotides, and the mechanism of the action of a polymerase upon binding to NDA, have been investigated through monitoring the 2AP steady-state and time-resolved fluorescence response. In addition to the studies based on the use of 2AP as a fluorescent probe, the spectroscopic properties of a new class of selenium-derivatised base analogues have been investigated experimentally and computationally.

This thesis is structured as follows. Chapter 2 presents the principles of fluorescence spectroscopy, background knowledge of DNA structure and the photophysical properties of 2AP-labelled DNA. The computational methods used to this work are also considered. Chapter 3 presents the experimental and computational methods.

Chapter 4 presents an investigation of the fluorescence response of 2AP in DNA single strands. The effects of base stacking and base pairing on the inter-base interactions that quench 2AP fluorescence are discussed. In Chapter 5 the fluorescence properties of 2AP in dinucleotides, the simplest DNA system, are reported. The base stacking interaction between 2AP and each natural base is explored and the influence of base dynamics on 2AP fluorescence is investigated by measurements at 77 K

Chapter 6 reports the use of 2AP time-resolved fluorescence to investigate of the interaction of an archael DNA-polymerase with primer-templates containing a deaminated base, uracil, in the template strand. It is shown that when the polymerase encounters a uracil 2 positions ahead of the primer-template junction, the enzyme unwinds the Watson-Crick base-paired primer-template. Previously, such unwinding had been postulated to be part of the proof-reading mechanism whereby the polymerase guards against the misincorporation of deaminated bases.

Chapter 7 presents the results of spectroscopic and computational studies on selenium-substituted base analogues, specifically nucleotides of thymine and guanine in which oxygen, at the 4 - and 6-positions, respectively, has been replaced with selenium. This substitution causes a dramatic shift of the absorption spectrum of the bases into the visible region. The potential of these base analogues as luminescent probes was investigated, but they do not show significant fluorescence or phosphorescence. The results of computational studies show that replacement of oxygen by selenium has almost no effect on the ground state structure of thymine and guanine, whereas it has big influence on the first excited-state structure in both cases. SCRF-PCM/CIS/6-311+G(d,p) calculations have been used to predict successfully the absorption spectra of the natural bases and the selenium-substituted

base analogues, and show that the selenium-induced shift in the absorption wavelength can be traced predominantly to decreases in the eigenvalues of unoccupied orbitals involved in the transitions.

Finally, Chapter 8 presents the conclusions drawn from this work.

Chapter 2

Theory and Background Knowledge

This chapter introduces the techniques of fluorescence spectroscopy and time-resolved fluorescence spectroscopy that are central to the studies presented in this thesis. Secondly, the essential properties of the molecules investigated, DNA and the fluorescent probe 2-aminopurine, the fluorescent analogue of the natural base adenine, are presented. Finally, the principles of the computational calculations are discussed.

2.1 Fluorescence Spectroscopy

2.1.1 The Principles of Fluorescence

The emission of light from any substance is called luminescence and occurs from electronically excited-states. Luminescence is divided into two categories, fluorescence and phosphorescence, as defined below. The excitation and de-excitation processes can be described by a simplified Jablonski diagram as shown in Figure 2.1. S_0, S_1, and S_2 represent the singlet ground, first excited, and second excited electronic states, respectively. T_1 stands for the triplet state in which the excited electron has the same spin orientation as the ground-state electron. There are a number of vibrational energy levels in each of these electronic energy levels, showed by 0, 1, 2, etc.

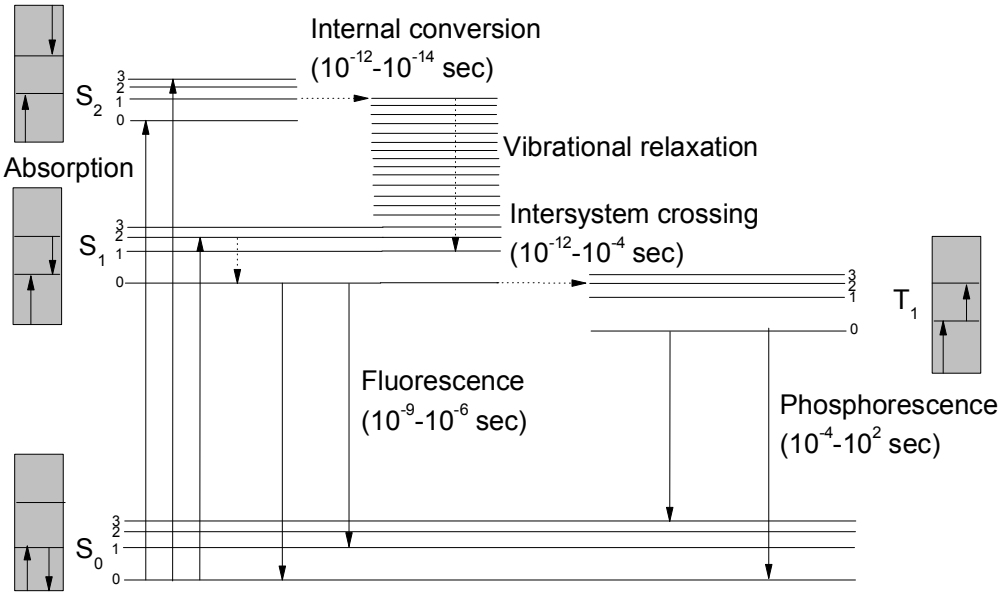


Figure 2.1 A Jablonski Diagram showing possible electronic transitions and deactivation processes. Radiative processes are shown by solid arrows and non-radiative processes by dotted arrows.

A molecule in the ground singlet state can absorb incoming light (within the ultra-violet and visible region) and be excited to higher energy states. This is illustrated by vertical lines between states such as from S_0 to S_1, or from S_0 to S_2. The transition occurs in about 10^{-15} seconds (s) that is too short for significant change of nuclear positions. If excitation of a molecule occurs to a higher electronic level than S_1, the molecule relaxes quickly (on the picosecond timescale) to the lowest vibrational level of S_1 by rapid internal conversion (IC) and vibrational relaxation (VR), shown by the dotted line. VR involves losing excess energy to the environment as heat. IC is an intramolecular process.

Possible decay processes of a molecule in the S_1 state include IC to S_0, inter-system crossing (ISC) to the first triplet state T_1 and emission of a photon to S_0, termed fluorescence. ISC involves a spin conversion of the excited electron with no electronic and vibrational energy loss. The emission of a photon from T_1 to S_0 is called phosphorescence. Since the excited electron in T_1 has the same spin orientation as the remaining electron in S_0, the electronic transition from T_1 to S_0 is forbidden,

resulting in a phosphorescence lifetime of milliseconds to seconds that is much longer than that of fluorescence.

Fluorescence and phosphorescence are radiative processes whereas VR, IC and ISC are non-radiative processes. Besides these processes, intermolecular processes, known as fluorescence quenching, result in excited-state relaxation. Fluorescence quenching is the decrease of the fluorescence intensity of the sample. Quenching from interaction of the excited fluorophore and other molecules is called dynamic or collisional quenching. Quenching through forming a complex between the ground-state fluorophore and other molecules is named static quenching. ISC induced by the external heavy atom effect, electron transfer and electronic energy transfer are the three major mechanisms for quenching.

Fluorescence excitation and emission wavelength (λ) and intensity (I) can be measured by steady-state spectroscopy. Wavelength is related to the energy difference between the ground - and excited-state. The relationship is shown by Equation 2.1.

Equation 2.1 $$E = h\nu = \frac{hc}{\lambda}$$

where h is Planck's constant, ν is frequency and c is the speed of light.

Fluorescence intensity describes the number of photons emitted per unit time and is related to the excited-state population. The population of excited fluorophore [M*] decays with first order kinetics after excitation with a pulse of light, as shown in Equation 2.2.

Equation 2.2 $$-\frac{d[M^*]}{dt} = (k_r + k_{nr})[M^*] = k_f[M^*]$$

where k_f is the sum of the rate constants including the radiative rate constant k_r and non-radiative rate constants k_{nr} (such as IC, ISC and quenching etc).

Integration of Equation 2.2 gives the dependence of fluorophore population [M*] on time t as shown in Equation 2.3.

Equation 2.3 $$[M^*](t) = [M^*]_0 \exp(-k_f t)$$

where $[M^*]_0$ is the concentration of excited-state fluorophore at t = 0.

Equation 2.4 shows the relationship between fluorescence intensity (I) and excited-state population [M*].

Equation 2.4 $$I(t) = k_r[M^*]$$

where k_r is the radiative rate constant.

Equation 2.5 expresses the relationship between fluorescence intensity (I) and time t by substituting equation 2.4 into equation 2.3.

Equation 2.5 $$I(t) = I_0 \exp(-k_f t)$$

where I_0 is the intensity at t = 0.

The fluorescence lifetime (τ) is the inverse of the total decay rate as shown in Equation 2.6.

Equation 2.6 $$\tau = \frac{1}{k_f}$$

By combining Equation 2.5 and 2.6, Equation 2.7 is generated from which fluorescence lifetime (τ) can be calculated from the measured decay of fluorescence. From this equation, the fluorescence lifetime (τ) is the time when the intensity decreases to 1/e of the intensity at t = 0.

Equation 2.7 $$I(t) = I_0 \exp(\frac{-t}{\tau})$$

The fluorescence quantum yield (Φ) and lifetime (τ) are the characteristic parameters of a fluorophore. The fluorescence quantum yield is the ratio of the number of photons emitted to the number absorbed; it is given by Equation 2.8. The maximum quantum yield is unity when the nonradiative decay rate is zero.

Equation 2.8 $$\Phi = \frac{k_r}{k_r + k_{nr}}$$

2.1.2 Time Correlated Single Photon Counting

Time-resolved measurements are widely used in the study of fluorescence and photophysics. The time-domain (TD) and frequency-domain (FD) methods are the two major methods for the time-resolved fluorescence measurements. The time-domain method, which is performed using time-correlated single photon counting (TCSPC), is employed in this work because of its ability to measure complex decay kinetics.

The principle of TCSPC can be explained simply as follows. The sample is excited with a laser pulse that is much shorter than the decay time. Experimental conditions are adjusted to ensure that no more than one photon is detected per excitation pulse. The time difference between the excitation laser pulse and a single emitted photon is detected and stored in a histogram. By exciting the sample many times, a histogram of single photon arrival times relative to the excitation laser pulse is obtained.

2.1.2.1 Fluorescence Decay Data and Analysis

Figure 2.2 shows a typical fluorescence decay curve collected by TCSPC. The instrument response, which is the response of the instrument to the excitation pulse, depends on the shape of the excitation pulse and the time response of the instrument.

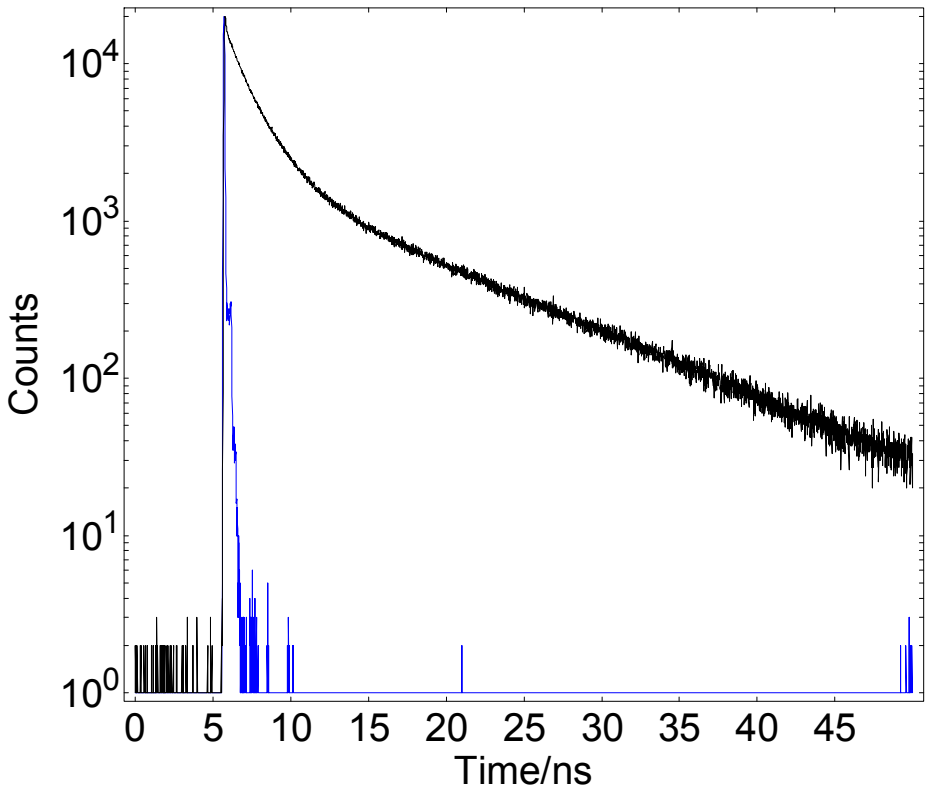


Figure 2.2 A typical fluorescence decay curve of 2AP-G dinucleotide in PBS buffer excited at 305 nm and collected at 380 nm by TCSPC. The instrument response function is blue and the measured sample decay is black.

The measured sample response function X(t) is the convolution integral of the instrument response function (IRF) R(t) and the sample decay function Y(t), as given by Equation 2.9.

Equation 2.9 $X(t) = R(t) \otimes Y(t)$

Intensity decays are typically fit to a multi-exponential model using Equation 2.10. The purpose of measuring the intensity decay is to obtain the amplitudes (α_i) and the decay times (τ_i).

Equation 2.10 $$I(t) = A + \sum_i \alpha_i \exp(-t / \tau_i)$$

where A represents the background signal, pre-exponential factor α_i represents the fractional amount of each distinct emitting species, and τ_i is the fluorescence lifetime of the i^{th} species.

Analysis of TCSPC data is generally carried out by the non-linear least squares (NLLS) method[1-3]. The purpose of the least squares fitting is to establish whether a given model is consistent with the decay data, and to determine the values of the variable for that model. The goodness-of-fit is defined by the χ^2 parameter (Equation 2.11).

Equation 2.11 $$\chi^2 = \sum_{i=1}^{n} \frac{[X(t_i) - Y(t_i)]^2}{X(t_i)}$$

where n is the number of data points, X(t) is the measured sample response function and Y(t) is the sample decay function.

Because χ^2 changes with the number of data points, the value of reduced χ^2 (χ_R^2) is normally used to judge the goodness-of-fit. The relationship between χ^2 and χ_R^2 is giving in Equation 2.12.

Equation 2.12 $$\chi_R^2 = \frac{\chi^2}{n - p}$$

where p is the number of variable parameters.

The value of (n-p) is approximately equal to n because the number of data points is far larger than the number of variable parameters, as shown in Equation 2.13.

Equation 2.13 $$\chi_R^2 \approx \frac{\chi^2}{n}$$

If the model fits the data perfectly, which means only random errors contribute to χ_R^2, then χ_R^2 is close to unity. If the model does not fit the data, χ_R^2 will be dramatically larger than unity. For the present work, on the basis of experience, values under 1.2 can be accepted as a good fit in general.

As well as using χ_R^2 to judge the goodness of fit, the weighted residuals D_i (Equation 2.14) should also be examined.

Equation 2.14 $$D_i = \frac{I(t_i) - Q(t_i)}{\sqrt{I(t_i)}}$$

where I (t_i) is the intensity of measured decay and Q (t_i) is the intensity of fitted decay.

For a good fit, the residuals are randomly distributed near zero as shown in Figure 2.3 (middle). Any deviation from randomness indicates a poor fit as shown in Figure 2.3 (bottom). This may come from instrument set-up problem or using the incorrect number of exponential terms in the model. Figure 2.3 shows an example of a good fit with four exponentials (top and middle) and a bad fit with three exponentials (bottom). It can be seen that three exponential fits show obviously non-random residuals with reduced χ^2 = 1.229. The fits with four exponentials does improve significantly with an extra exponential, and the reduced χ^2 = 1.020. Therefore, four-exponential fits are necessary for this decay curve.

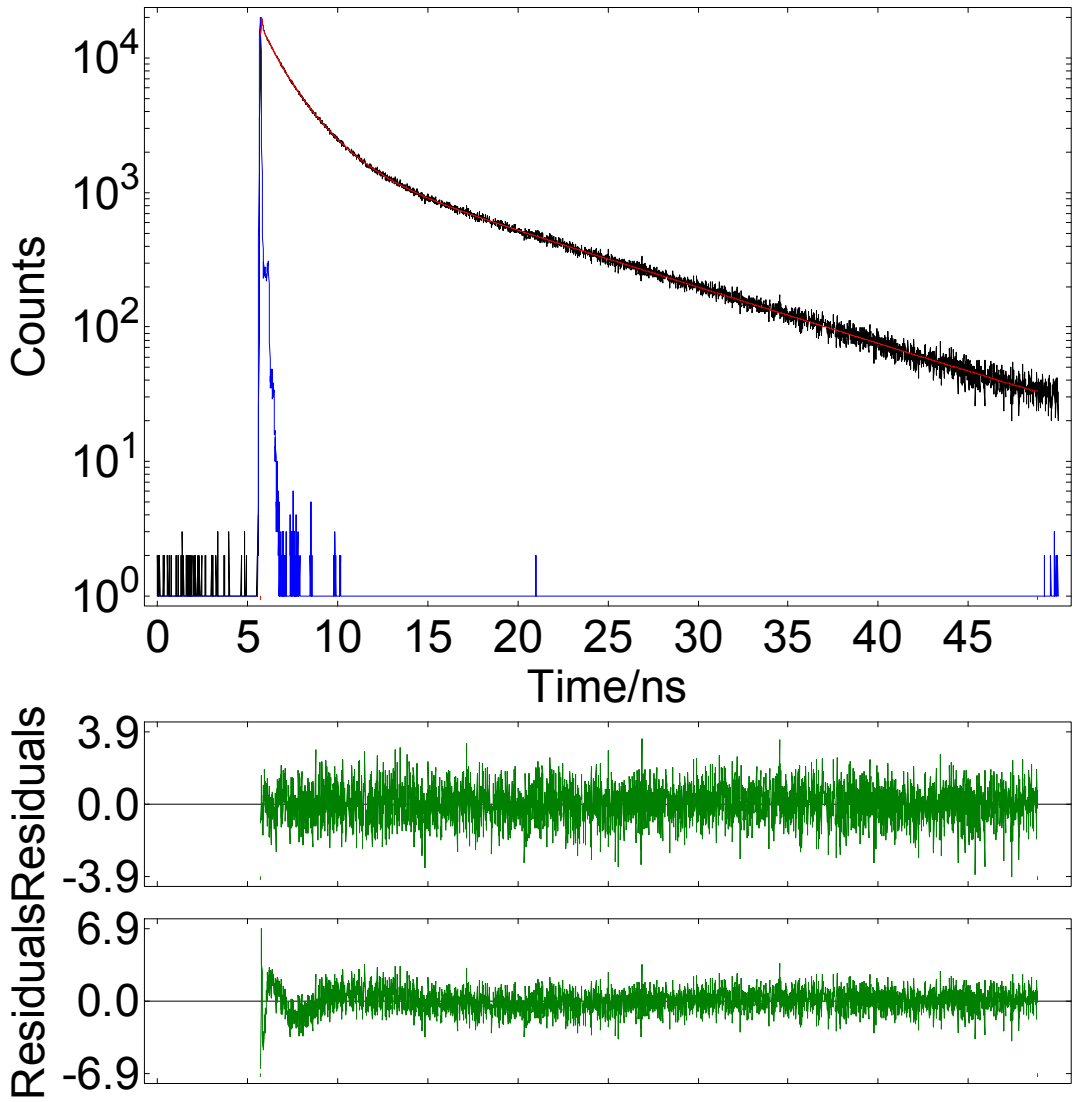


Figure 2.3 An example of a good fit of the decay curve (shown in Figure 2.2) with four exponentials (top and middle) and a bad fit with three exponentials (bottom). The IRF is blue, the measured sample decay is black, the fitted sample decay is red and the residuals are green. When the decay was fit with four exponentials, by iterative reconvolution with a separately measured IRF, between the 470th and the 4000th time channel, with a background level as 0.557 count, the resulting lifetimes calculated for the red function are 0.06 ns, 0.60 ns, 1.8 ns and 10.3 ns and the reduced χ^2 = 1.020. When the decay was fit with three exponentials, under the same condition as for the four exponentials, the resulting lifetimes calculated are 0.16 ns, 1.6 ns and 10.2 ns and the reduced χ^2 = 1.229.

Fitting the amplitudes and decay times in a multi-exponential decay has difficulties because of correlation between the variable parameters. Measuring several sets of decay data at different wavelengths and doing global analysis improve the resolution of correlated parameters[4-9]. In the present work, typically global analysis was carried

out on three fluorescence decays recording at different emission wavelengths. This procedure involves simultaneous fitting of the fluorescence decays with common lifetimes and is based on the assumption that the fluorescence lifetimes are independent of the emission wavelength, but their amplitudes are wavelength-dependent.

The fractional contribution (f_i) of each decay time to the steady-state intensity can be calculated using α_i and τ_i (Equation 2.15).

Equation 2.15 $$f_i = \frac{\alpha_i \tau_i}{\sum_i \alpha_i \tau_i}$$

The average lifetime of the sample is calculated using Equation 2.16.

Equation 2.16 $$\langle \tau \rangle = \frac{\sum_i \alpha_i \tau_i}{\sum \alpha_i}$$

2.2 Biological Context

2.2.1 DNA Structure

DNA is the acronym for deoxyribonucleic acid, a polymeric molecule that is formed from deoxyribonucleotide monomers. Each nucleotide is grouped into three parts, which are base, sugar and phosphate. The monomers are joined together to form the backbone of DNA by the repeating sugar-phosphate unit through a 3'-5' phosphodiester bond. There are four bases connected to the DNA backbone which are either purine, adenine (A) and guanine (G), or pyrimidine, thymine (T) and cytosine (C), bases. The structures of the DNA backbone and the bases are illustrated in Figure 2.4. Purine bases bind through the atom N_9 and pyrimidines through the atom N_1 to the sugar C_1.

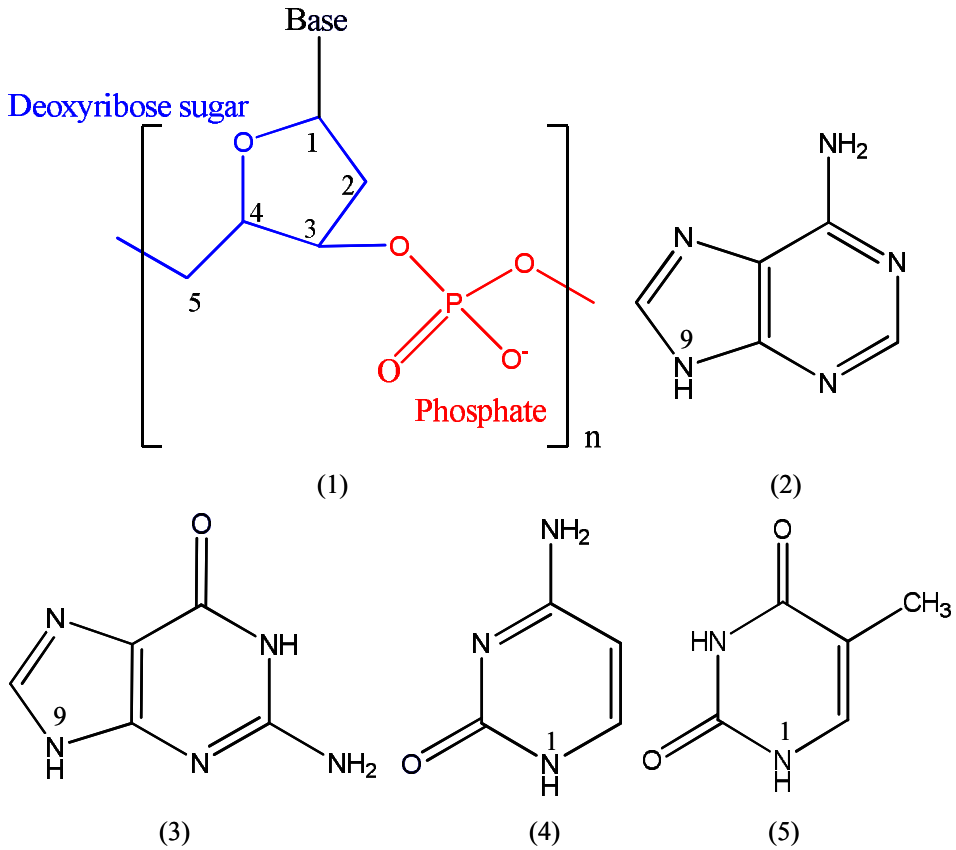


Figure 2.4 Structures and numbering schemes for the (1) sugar-phosphate backbone, (2) adenine, (3) guanine, (4) cytosine and (5) thymine.

The DNA structure has two important features. One is that the bases are buried at the centre of the helix due to their hydrophobic nature compared with the DNA backbone. The other is that the form of a typical DNA is a 'double helix'. It is made by two single strands spiraling around each other. These two strands coil in opposite directions. Pairs of bases interact to maintain the famous double helical structure stable through hydrogen bonding. The most stable base pairs, as determined by Watson and Crick[10], are A with T and G with C. The hydrogen bonded base pairs and double helical structure of DNA are shown in Figure 2.5.

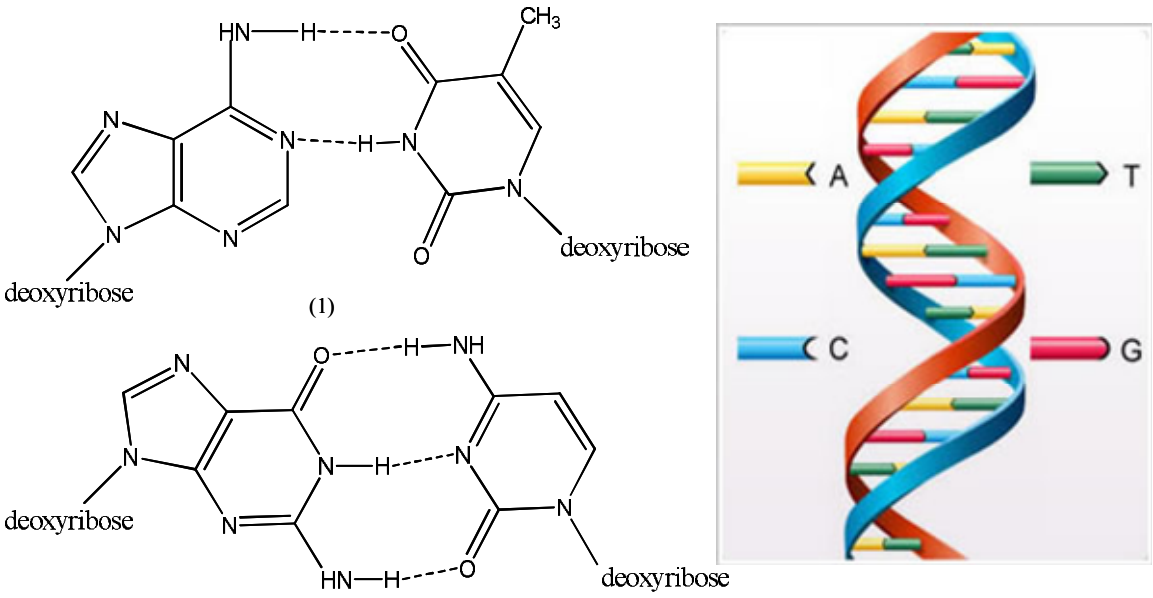


Figure 2.5 The hydrogen bonding of Watson-Crick base pairs, (1) A and T, (2) G and C, and (3) double helical structure of DNA (taken from the website www.eurofinsforensics.co.uk).

In real world, different types of DNA double helixes have different number of bases per turn of helix. 'A' form of DNA has 11 bases per turn, 'B' form of DNA has 10 and 'Z' form of DNA contains 12 bases per turn. The 'A' and 'B' forms of DNA are right-handed, whereas the 'Z' form of DNA is left-handed. The structures of A, B and Z forms of DNA are represented in Figure 2.6.

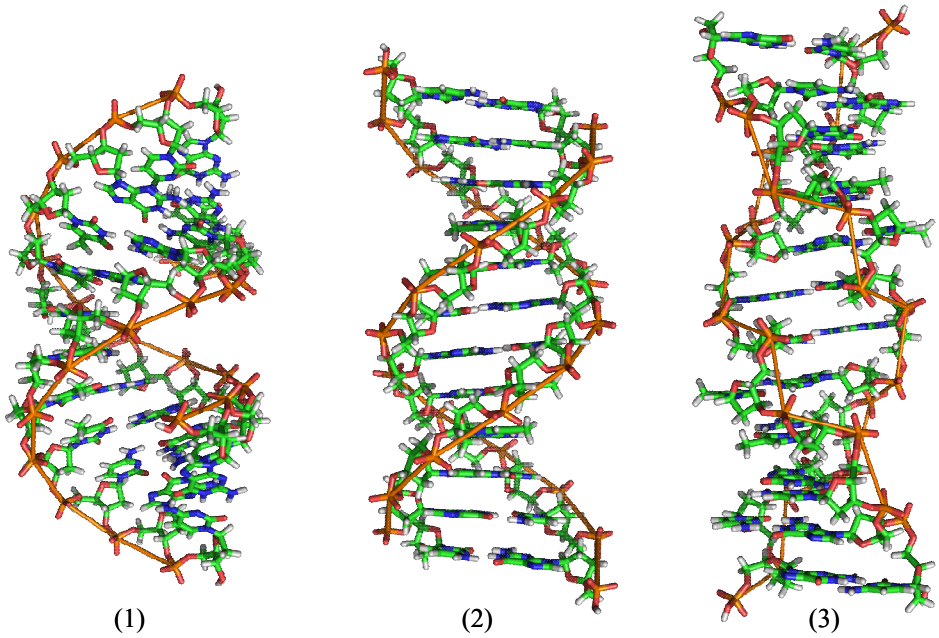


Figure 2.6 The structures of (1) A form, (2) B form, and (3) C form of DNA. Reproduce from website en.wikipedia with uploader of Richard Wheeler.

In the most common B-form DNA, interstrand separation between pair of bases is 3.4 Å, as shown in Figure 2.7. The diameter of the helix is 20 Å. The helical pitch is 35.7 Å. The twist angle rotation between the neighbouring bases is 36 °, therefore in theory roughly 10 bases are needed to represent a complete turn of B-form DNA.

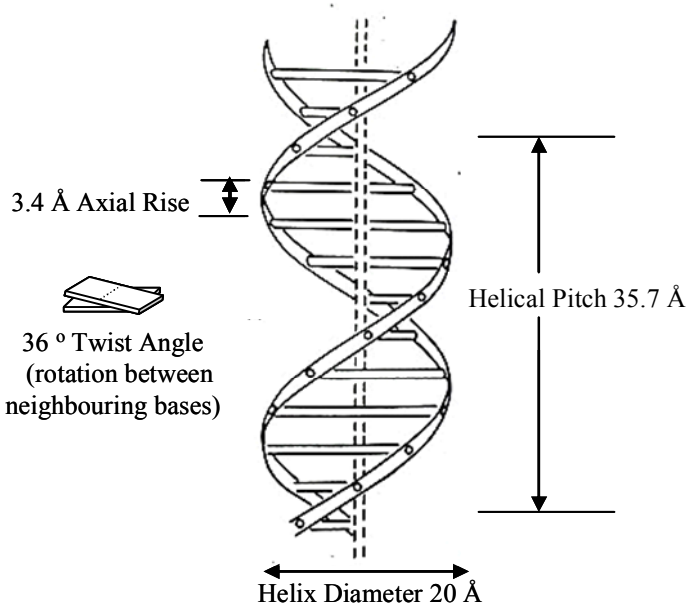


Figure 2.7 The structure of the B-form DNA. The dashed vertical line is the long axis of the helix, and the ribbon-form represents the DNA backbone. The horizontal double lines connecting the two ribbons are the base pairs.

The DNA molecule is dynamic with its base pairs breaking and reforming on the millisecond timescale. There are also three motions of any base relative to its neighbour, which are slide, twist and roll. Slide describes the relative sliding of neighbouring base along their long axes. Twist describes a rotation between any two neighbouring bases around the twist axis, which is the centre of these two bases. Roll describes the rotation open between neighbouring bases along their long axes. Figure 2.8 shows these three motions.

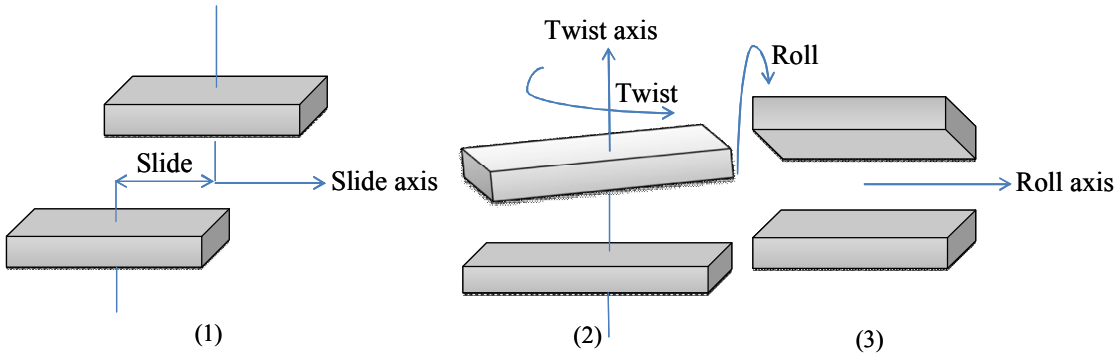


Figure 2.8 The (1) slide, (2) twist, and (3) roll motion between neighbouring bases in DNA.

The DNA sequence determines the three-dimensional structure and many biochemical properties of DNA. The sugar-phosphate backbone is independent of the DNA sequence; therefore the effect of variation of DNA sequence on the three-dimensional structural properties of DNA relies on the subtle differences in the covalent structure of the bases[11]. In DNA, adjacent bases and base pairs form non-covalent π-stacking interactions that are mainly determined by van der Waals and electrostatic energies[11,12]. The van der Waals energy includes the dispersion and repulsion energies. The dispersion energy is an induced dipole-induced dipole interaction that relates to the intermolecular surface contacting area[11]. The repulsion energy only exists when the atomic distances are shorter than their van der Waals contact distance[11]. The electrostatic energy arises from Coulombic interactions between the positive and negative electron densities across the aromatic bases and depends on the molecular charge distributions and π-electron density on the aromatic rings[11]. Each base has a distinct charge distribution due to different electron-negativities of atoms in its σ-framework. Furthermore, there is negatively

charged π-electron density below and above the aromatic ring and positively charged σ-framework in all the aromatic bases[11].

2.2.2 2-Aminopurine: a Fluorescent Probe for DNA

It is necessary to use extrinsic fluorescent probes to monitor changes in DNA conformation and understand interactions of DNA with enzymes because the natural bases of DNA are non-fluorescent due to their extraordinarily low quantum yields (Φ = 0.5-1.2 x 10^{-4} for nucleotides)[13].

2-Aminopurine (2AP) has been used for nearly thirty years as a probe of nucleic acid properties. 2AP is a fluorescent analogue of A and its structure is similar to A, as shown in Figure 2.9. The only difference between 2AP and A is the location of the exocyclic amine group, which is at C_2 for 2AP and C_6 for A. The quantum yield for 2AP riboside (2APr) is about 0.68[14], far greater than the natural bases. When 2AP is in the DNA helix, it forms Watson-Crick base pairs with T through hydrogen bonding which maintains the structure of double helix[15]. The maximum intensity of the 2AP excitation spectrum is 305 nm and maximum emission occurs at about 370 nm. The excitation spectrum of 2AP is red-shifted compared with the DNA natural bases, which have their maximum absorption intensity at approximately 260 nm, therefore, selective excitation and detection of 2AP can be achieved[16]. Furthermore, the fluorescence properties of 2AP are sensitive to local and global molecular environment[17,18]. Hence, 2AP is a suitable quasi-intrinsic fluorescent probe to monitor DNA structure and dynamics[14,19-22]. It also can be used to investigate DNA or RNA and protein interactions[23-30].

Adenine tautomerises by transferring a proton between N_7 and N_9. 2AP is similar to this, showing 7H (40% of the emitting population) and 9H (60% of the emitting population) tautomeric forms in aqueous solution at room temperature[31]. In 2APr aqueous solution, the replacement of the 9H proton by the ribose group prevents the

formation of 9H tautomer. The structure of 2AP, A, 2APr and 7H and 9H tautomers of A are shown in Figure 2.9.

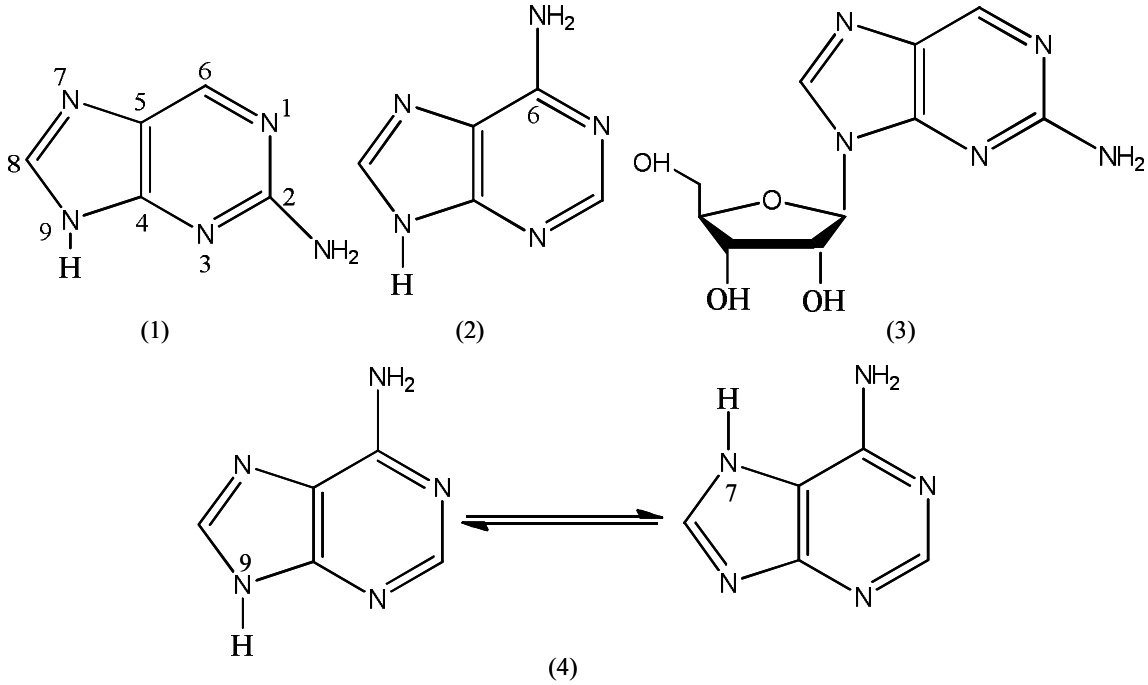


Figure 2.9 The structures of (1) 2AP (9-H tautomer), (2) A (9-H tautomer), (3) 2APr, and (4) the 9H and 7H tautomers of A.

2.2.2.1 Long Wavelength Emission of 2AP in DNA Duplex

When 2AP is incorporated into DNA and excited at shorter wavelengths (≤ 340 nm), it generates an emission band with maximum intensity at 370 nm that is termed the 'normal' fluorescent emission spectrum. The maximum intensity of the 2AP-labelled DNA excitation spectrum with emission wavelength 370 nm is about 310 nm. Therefore, the 'normal' fluorescent excitation and emission spectra of 2AP-labelled DNA are almost identical to those of free 2AP in solution.

2AP-labelled DNA has a long wavelength emission with maximum intensity at about 450 nm, which has been pointed out by several research groups. The 450 nm emission band is termed 'long wavelength' emission of 2AP-labelled DNA as

compared with the 'normal' emission band. As early as in 1990, Guest *et al*.[32] observed that the emission spectra of 2AP labelled 7-mer oligonucleotide duplexes exhibited a peak at 380 nm and the tail of the emission spectra broadened to the long wavelength when the duplexes were excited at 320 nm. They attributed the long wavelength emission to exciplex formed between 2AP and the nearest neighbouring G based on the sequence of the DNA duplexes.

Rist *et al*.[33] investigated two different types of 2AP-labelled DNA duplexes. One contained pairs of adjacent 2AP molecules and the other contained some single 2AP molecules that were located between natural DNA bases. When the two 2AP-labelled DNA duplexes were excited at 320 nm, the typical 2AP fluorescence with a maximum about 370 nm was generated. There was an additional emission band with maximum intensity at 440 nm as well as 370 nm emission when the 2AP-labelled DNA duplexes were excited at 343 nm. The intensity of the long wavelength emission band for duplexes containing the single 2AP molecules was only one quarter of that with adjacent 2AP molecules. The 440 nm emission band became more intense and the normal 370 nm emission band nearly disappeared when they were excited at 350 nm or longer. At first, Rist *et al*.[33] thought this long wavelength emission came from 2AP excimers, but the excitation spectrum with an emission wavelength of 440 nm had two peaks, one was at 310 nm, which was normal for 2AP, the other was at 370 nm. Rist *et al.* concluded that there was a delocalised species, which came from ground-state interactions between adjacent 2AP molecules and can be excited separately.

Neely *et al*.[34] investigated the steady-state fluorescence spectra of single-crystal 2AP. The emission spectrum with a peak at approximately 370 nm was excited at 300 nm or less. This was similar to that of 2AP in aqueous solution. The red-shifted emission spectrum with a peak at 420 nm was excited at longer wavelengths. The results indicated that single-crystal 2AP has two ground-state species with different electronic structures giving rise to two distinct emission bands with their own excitation spectra. The intensity of the longer wavelength emission is about five

times more intense than that of the 'normal' wavelength emission. When the crystalline 2AP was dissolved in water, the 'normal' emission wavelength remained and the long wavelength emission no longer existed. Neely *et al.* attributed the long wavelength emission to dimer - or trimer-like ground-state species resulting from the π-stacking interaction between 2AP and its neighbouring 2AP molecules. They also found the long wavelength emission existed in a 2AP-labelled DNA duplex, with emission maximum at approximately 430 nm when the DNA duplex was excited at 360 nm[34]. However, the intensity of this emission was only one quarter of that of the 'normal' 370 nm emission band, which was different to the trend of single-crystal 2AP. It was suggested the difference might be due to the different π-stacked structures between the duplex and the 2AP crystal. 2AP molecules formed the highly stacked conformation in the crystal lattice, whereas 2AP and its neighbouring bases have the reduced π-stacked conformation due to the rotational offset (~36 °, as shown in Figure 2.7) between adjacent bases. The rotational offset needed to be decreased to a certain degree in order to facilitate formation of the dimer-like species in DNA duplexes.

Bonnist *et al.*[22] recently reported a systematic study of the long wavelength emission band of 2AP-labelled DNA, in a variety of duplexes and some single strands. This band, with maximum emission at 450 nm is observed at excitation wavelengths ≥ 350 nm. This is attributed to a heterodimer species formed when 2AP is closely π-stacked with one of its neighbouring natural bases. The heterodimer has its own excitation spectrum with maximum intensity at 360 nm and is a ground-state species. The intensity of the long wavelength emission is much weaker than that of the 'normal' emission. Also, the intensity of the long wavelength emission in DNA duplexes is affected less by sequence context than that of the 'normal' wavelength emission.

2.2.2.2 Quenching of 2AP Fluorescence in DNA Duplex

The fluorescence intensity of free 2AP in solution is high whereas that of 2AP inside DNA decreases significantly. The most likely quenching mechanism of 2AP fluorescence in DNA is electron transfer (ET) through π-stacking interactions between adjacent bases and base pairs[35-39]. The most efficient and favourable ET is from G to 2AP excited-state (2AP*), due to the largest redox potential difference between G (1.29 V versus NHE)[40] and 2AP* (1.5 V versus NHE)[35].

O'Neil *et al.*[41] showed the first direct evidence for ET from G to 2AP* in DNA through measuring HPLC traces of (2AP)AAA^{CP}G (cyclopropylguanine) duplexes following 325 nm irradiation and enzymatic digestion to the nucleosides. Previously, it had been shown that the oxidation potential, base pairing and stacking properties of ^{CP}G are similar to those of G[41,42]. Therefore, ^{CP}G could be used to trap the intermediate oxidised G radical cation before redistribution of charge through back ET (BET). The hole trap is based on the rapid opening of the cyclopropyl ring in ^{CP}G radical cation[42]. The mechanism of hole trapping is illustrated in Figure 2.10.

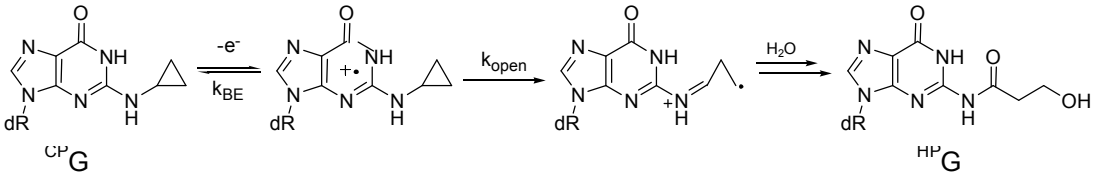


Figure 2.10 The mechanism of hole trapping by ^{CP}G. Reproduced from reference 42.

The HPLC traces before and after irradiation of (2AP)AAA^{CP}G are shown in Figure 2.11. It can be seen that ^{CP}G undergoes the photoinduced decomposition, suggesting the hole transferred to ^{CP}G from 2AP* can be trapped before BET. When the 2AP in DNA duplexes is replaced by A, the loss of ^{CP}G is not observed following 325 nm irradiation. Also, when ^{CP}G is replaced by G in the above duplexes, there is no photoinduced damage detected[41]. Therefore, O'Neil *et al.*[41] suggested BET is considerably faster than trapping of the G radical cation by water and/or oxygen.

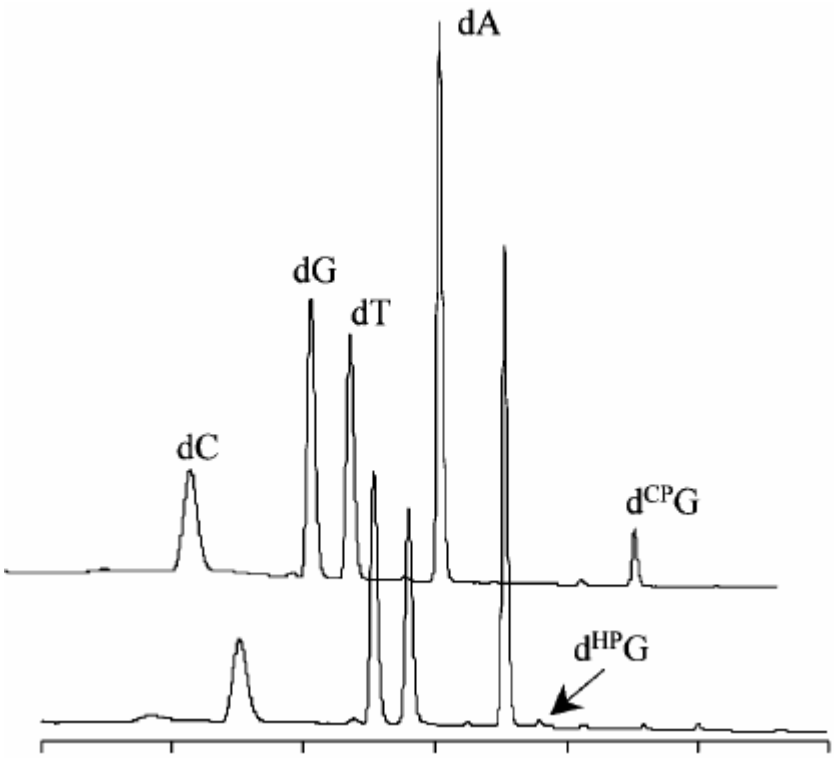


Figure 2.11 HPLC traces (monitored at 260 nm) before and after 325 nm irradiation (30 minutes) of (2AP)AAA^{CP}G duplexes. Reproduced from reference 41.

Kelley *et al*.[35] explored the quenching of 2AP fluorescence by the π-stack-mediated ET between the natural bases and 2AP*. They studied the fluorescence steady-state spectra and lifetimes of a series of 2AP-labelled 12-base pair duplexes. Their results showed that ET quenching can happen at longer donor (G) - acceptor (2AP*) separation, but the quenching efficiency was lower than that at shorter distance. The quantum yield of 2AP-labelled DNA duplex with 2AP and G at adjacent positions was 0.0005, whereas that with 2AP and G separated by three other natural bases was almost eight times greater. With increasing the distance between 2AP and G in the above duplexes, the magnitude of the shortest fluorescence lifetime component of 2AP increased, suggesting the efficient quenching becoming considerably weaker[35]. The intrastrand ET happens on a timescale less than 100 ps. This ET quenching is classified as static quenching, because it is much faster than the time scale of other 2AP* deexcitation processes[35]. The ET rate is at least 10^{10} s^{-1} [35].The reaction kinetics and distance dependences of ET processes were affected by the extent of π-stacking interactions between neighbouring bases and base pairs.

Later, Wan *et al.*[36] investigated ET between 2AP and natural bases in DNA by using femtosecond techniques. The transient absorption and fluorescence up-conversion of fourteen 2AP-labelled DNA duplexes were measured. When the bridging bases between 2AP* and G were A, long distance ET between 2AP* and G could occur easily. When the bridging bases were T and C, long distance ET between 2AP* and G was much less efficient. The ET efficiency between 2AP* and G was strongly distance dependence. However, over three bridging bases, this distance dependent property of ET was not significant.

O'Neil and Barton[37] studied the impact of directionality on ET between 2AP* and DNA bases. They found that the ET was much faster in the 3'-5' direction than in the 5'-3' direction. Their results also showed that the relative quantum yield of 2AP was dramatically sensitive to subtle structural variations in the DNA duplexes. The later work of Shao *et al.*[43] also proved that ET was significantly sensitive to DNA sequence and structure.

The driving force for ET between 2AP* and G in DNA duplex is the different reduction potentials between them. The reduction potentials of the other three natural bases A, C and T are 1.42, 1.6 and 1.7 V versus NHE, respectively[40]. Therefore, there are also differences of reduction potentials between 2AP* and A, or C or T as well as between 2AP* and G, indicating the ET probably can happen between 2AP* and A, or C or T. Using the transient absorption and fluorescence up-conversion techniques, Wan *et al.*[36] measured the fluorescence lifetimes of 2AP-labelled DNA duplexes with the central sequences being IApXI from 5' to 3' direction, where I is inosine, Ap is 2AP and X stands for the four natural bases. Since the reduction potential of I is the same to that of 2AP*[35], I is redox inactive and can be used as a reference for ET reactions and a probe for the non-redox interactions between 2AP* and the natural bases. The shortest lifetime component of IApGI was 19 ps that was almost identical to that of IApTI (20 ps). This was an indication that ET could happen through both oxidative and reductive reactions of the nucleotides with 2AP*. 2AP* could have reductive ET with G and A and oxidative ET with T and C in DNA. However, the

most efficient charge transfer was between G and 2AP* due to the largest difference of reduction potentials, as discussed above.

Fiebig *et al.*[44] studied the 2AP in four DNA complexes with mononucleotides of A, C ,G and T triphosphate to investigate the importance of ET using time-resolved transient absorption and fluorescence up-conversion methods. Their results also suggested both ET and hole transfer (HT) could happen when 2AP was used as a probe base in DNA. ET occurred from 2AP* to C and T, whereas HT occurred from 2AP* to G and A. They also proposed that both HT and ET from 2AP* to A could happen, suggesting its ambivalent redox property. However, whether HT or ET occurred from 2AP* to A depended on the local binding sites and microscopic environment. The 2AP-labelled DNA complexes had two fluorescence lifetimes: one was in ps range and the other one was in ns. If the ps (the shortest) fluorescence lifetime of 2AP-labelled DNA complexes involves ET between 2AP* and natural bases, it can be used as a method to obtain the efficiency of ET between 2AP* and the four natural bases. The shortest fluorescence lifetime of 2AP-labelled DNA complexes measured by using transient absorption technique was 19 ps (with G), 20 ps (with T), 47 ps (with C) and 350 ps (with A). Therefore, the efficiency of ET between 2AP* and natural bases from high to low was deduced to be G > T > C > A, which as completely consistent with the difference in reduction potentials between 2AP* and the natural bases.

2AP-labelled 35-mer DNA assemblies have also been studied through femtosecond spectroscopy by O'Neill *et al.*[39] in order to investigate the rate and character of ET. They proposed a conformational gating model for the ET in 2AP-labelled DNA duplexes. During the time course of excitation, a DNA solution was made of assemblies with a heterogeneous distribution of base conformations. The donor, acceptor and the bridge bases (optional) in some of the assemblies were in a conformation upon excitation that facilitated immediate ET (direct ET). This conformation was named ET-active and represented by the fastest fluorescence lifetime of 2AP-labelled DNA duplexes. Other assemblies upon excitation underwent base motions that resulted in the reorganization of the conformations during the

lifetime of 2AP*. Some assemblies adopted the ET-active conformation in which 2AP* decayed through the gated ET by initial reorganizational motion and had a slower ns fluorescence lifetime. Since the DNA base motions behaved as the gate, DNA dynamics had a significant effect on the population of the gated ET assemblies. There are also some assemblies not assuming ET-active conformation in which 2AP* decayed through non-ET pathways. When the donor and acceptor were in direct contact (ApG duplex), ET happened in 10 ps due to the high probability of being in an ET active conformation. When the donor and acceptor were separated by some other three natural bases (such as ApAG), ET occurred in about 100 ps due to undergoing some reorganizational motion. The efficiency of bridge-mediated ET could be enhanced with increasing base dynamics. However, the base dynamics had little influence on the direct contact ET. They also measured the fluorescence lifetimes of the 35-mer 2AP-labelled DNA single strands for comparison. The fast lifetime component was 20 ps in ApG single strand that was twice than that in the corresponding duplex. The ET rate decreased 50% in ApG single strand than in duplex. The fast lifetime component in ApAG was 420 ps that was significant longer than the corresponding one in duplex. Therefore, stacking was extremely important for efficient ET.

Subsequent studies by O'Neill and Barton[45] add further weight to the conformational gating model for the ET in DNA. They studied the temperature-dependent rate constant for charge transfer between 2AP* (acceptor) and G (donor) in 35-mers DNA containing A bridges $Ap(A)_nG$ (n = 0-9) and mixed bridges ApAAIAG and ApATATG by using femtosecond spectroscopy. It is the further investigation based on the above paper[39] and the bridging bases in 35-mer DNA are increased in this paper. They found the yield of charge transfer in DNA duplex raised with increasing the temperature. Furthermore, at high temperate the charge transfer becomes less dependent on the distance. They also showed that the yield of charge transfer exhibited a periodic oscillatory behaviour instead of simple exponential distance dependence at all temperatures. The results suggested charge was delocalized over domains, transient π-orbitals extended over 4-5 base pairs and charge transfer

occurred through hopping among these domains. They also confirmed that base dynamics was critical to DNA ET.

The rate constants and yields of ET are governed by three major aspects: the redox potential difference between the electron donor and acceptor, π-stacking interactions and base dynamics (motions)[35-37,39,45-47].

2.2.2.3 The 'Normal' and Long Wavelength Emission Decays of 2AP-labelled DNA Duplex

The fluorescence decay of 2APr in aqueous solution is mono-exponential with a lifetime of 10.6 ns[48]. The 'normal' fluorescence emission decay of 2AP monomer in DNA is multi-exponential and can be typically described by four lifetimes of < 100 ps, approximately 0.5 ns, 2 ns and 10 ns[20,28,32,49,50]. The distinct lifetimes represent different conformations of the DNA duplex. The shortest lifetime represents a conformation shown in Figure 2.12 (1), in which 2AP and one of its neighbouring bases are highly stacked and ET is favourable. The longest lifetime of 2AP in DNA is comparable to the lifetime of free 2APr in solution, therefore it corresponds to a conformation in which 2AP is extrahelical and free from π-stacking interaction, as shown in Figure 2.12 (3). The two intermediate lifetimes represent conformations 'intermediate' between the two extremes in which 2AP does not form perfect π-stacking interactions with its neighbouring bases and ET is less favourable[48]. Three imperfect stacked conformations are shown in Figure 2.12 (2). Each lifetime component of 2AP in DNA can be considered to represent a group of sub-conformations of DNA duplex in which 2AP undergoes the similar quenching interactions and hence presents a similar lifetime. Typically, about 70% of the DNA duplex is in a highly stacked conformation, whereas only 5% of the DNA duplex experiences an extrahelical environment. The rest of the DNA duplex (~25%) is in the intermediate conformations. Therefore, the majority of 2AP in DNA is intrahelical and forms stacked conformation with its neighbouring natural bases.

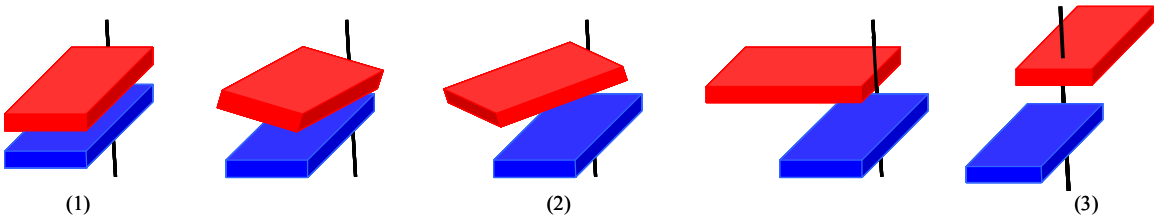


Figure 2.12 The structures of (1) highly stacked conformation, (2) imperfect stacked conformation and (3) free from the stacked conformation of DNA duplexes. 2AP is red rectangular and the other bases are blue rectangular. Reproduced from reference 51.

The long wavelength fluorescence emission decay of 2AP heterodimer in DNA is also multi-exponential and described by four lifetimes of ~0.17 ns, ~1 ns, 3.6 ns and ~12 ns[22]. Like 2AP monomer, 2AP heterodimer also exists therein a variety of conformational states with different quenching rates, described by the 'stacked', 'intermediate' and 'extrahelical' classifications. Broadly speaking, the monomer and heterodimer appear have a comparable quenching mechanism. The shortest lifetime of 2AP heterodimer is considerably longer than that of 2AP monomer in DNA, indicating that the very rapid quenching process is not present. This is probably because the 2AP dimer is geometrically constrained in the DNA conformation. The population of the 'stacked' conformation of 2AP heterodimer species decreases to ~50% compared with that of 2AP monomer. However, it is still the dominant species in DNA. The population of the 'extrahelical' species of 2AP dimer is almost constant to that of 2AP monomer. The reduced population of the 'stacked' conformation is transferred to the 'intermediate' conformations.

2.3 Theoretical Calculations

2.3.1 Molecular Quantum Mechanics

The foundation of quantum mechanics is the Schrödinger equation (Equation 2.17). The aim of the molecular quantum mechanics is to solve the Schrödinger equation to predict the electronic structures and properties of atoms and molecules.

Equation 2.17 $\hat{H}\Psi = E\Psi$

where $\hat{H}$ is the Hamiltonian operator and represents the total energy of a system, including kinetic energy (T) and potential energy (V). E is the energy, an eigenvalue of the operator $\hat{H}$. Ψ is the wavefunction of the system that includes information about all the properties of the system. The relationship between Hamiltonian operator and kinetic ($\hat{T}$) and potential energy operator ($\hat{V}$) is shown in Equation 2.18.

Equation 2.18 $\hat{H} = \hat{T} + \hat{V}$

In order to solve the Schrödinger equation, the Born-Oppenheimer approximation is used to simplify the general molecular problem by separating nuclear and electronic motions. Because the masses of the nuclei and electrons have enormous difference, nucleic can be treated as fixed and electrons are considered to move instantaneously in response to changes in nuclear position. Therefore, the Schrödinger equation for the electrons in the static electric potential resulting from nuclei is written in Equation 2.19.

Equation 2.19 $\hat{H}\psi(r,R) = E(R)\psi(r,R)$

where ψ is the wavefunction for the electron based on the position of the nucleus, r is the distance between the electron and nucleus, R is the position of nucleus and E(R) is the electronic energy depending on the nuclear coordinates.

For a fixed position of the nucleus, the Hamiltonian operator is shown in Equation 2.20. The first term on the right hand side of the equation is the kinetic energy of electrons, the second term is the electron and nucleus attraction potential energy and the third term is the electron-electron repulsion potential energy. The nucleus-nucleus repulsion potential energy, which is omitted in this equation, is added as a classical term at the end of the calculation.

Equation 2.20 $$\hat{H} = -\frac{\hbar^2}{2m_e}\sum_i^n \nabla_i^2 - \sum_i^n \sum_I^N \frac{Z_I e^2}{4\pi\varepsilon_0 r_{Ii}} + \frac{1}{2}\sum_{i,j}^n \frac{e^2}{4\pi\varepsilon_0 r_{ij}}$$

where $\hbar = h / 2\pi$ is a convenient modification of Planck's constant, m_e is the mass of an electron, $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, the three mathematical operators $\sum_i^n$, $\sum_i^n \sum_I^N$ and $\sum_{i,j}^n$ are the summations over all the electrons and nuclei, Z is the atomic number (number of protons in nucleus), e is the charge on an electron, ε_0 is vacuum permittivity, r_{Ii} is the distance between electron and nucleus and r_{ij} is the distance between two electrons.

There are three major approaches to solve the above equation: *ab initio* methods, density functional theory and semiempirical methods. All of the theories described below belong to the first two approaches.

The crucial difficulty of solving the Schrödinger equation arises from the third term in Equation 2.20, the electron-electron repulsion potential energy. Dealing with this problem is discussed below.

2.3.2 The Hartree-Fock Self-consistent Field Approach

The changes in motion or position of electrons caused by the other electrons are termed as electron correlation. In other words, correlation means the instantaneous motion of one electron affects the motions of all the other electrons in the system due to their mutual Coulombic repulsion.

According to the Pauli principle, the total wavefunction (including spin) must be antisymmetric with respect to the interchange of any pair of electrons. As a consequence, electrons with parallel spins tend to avoid one another, which is known as electron spin correlation. This correlation is named exchange correlation and is the

correlation resulting from the electron pairs of the same spin. In order to add this spin correlation effect to the wavefunction, a spinorbital that is a combination of an orbital wavefuntion and a spin function, is introduced. The wavefunction that satisfies the Pauli principle is written as the Slater determinant.

In the Hartree-Fock (HF) method, electron correlation is treated in an average way, which means each electron is treated as moving in the average electric field of the nucleic and all the other n-1 electrons. The neglect of correlation means that bond lengths predicted by HF tend to be shorter than the ones obtained from experiment, and therefore the bond strengths are over estimated in HF method.

The HF equations are solved by using the self-consistent field (SCF) method, in which an initial set of spinorbitals is generated. Then the HF equations are solved in order to obtain a new set of spinorbitals, which in turn generate the new HF equation, and so on. This procedure is repeated until the new set of spinorbitals meet a convergence criterion to some specified tolerance. Self-consistent means the new set of spinorbitals does not change too much compared with the set of spinorbitals obtained in last cycle.

2.3.3 Post-HF Approaches

As described above, HF theory treats the electron correlation in an average way and the instantaneous repulsion between electrons is neglected. However, electron correlation is important in order to obtain the accurate results compared with experiment. For quantum chemical methods to reach a useful level of chemical accuracy, any method that attempts to include this electron correlation is called post-HF method. Configuration Interaction (CI) and Møller-Plesset (MP) many-body perturbation theories are two widely used methods, and are introduced below.

2.3.3.1 Configuration Interaction

HF introduces the concepts of occupied and unoccupied orbitals. The occupied orbitals are filled in the HF ground-state wavefunctions. The HF wavefunctions generate the lowest energy configuration of electrons among all the orbitals. The unoccupied orbitals are named as virtual orbitals and are higher energy than the occupied orbitals. Electrons can be permutated from an occupied orbital to a virtual orbital, which efficiently enlarges the space of electrons movement, as shown in Figure 2.13.

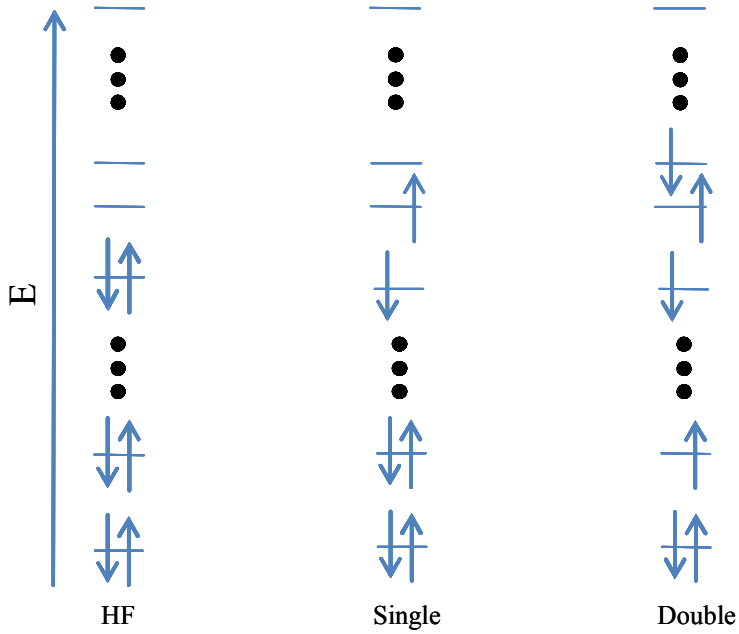


Figure 2.13 An example of electrons permutated from HF level (left) to occupy the higher energy virtual ortibals (the middle and right).

CI uses the HF ground-state and includes contributions from higher excited determinants, which allows electrons to obtain larger separation. CI is a method where the wavefunction is expressed as a linear combination of determinants. It includes the electron correlation that is neglected in the HF method. The exact electronic wavefunction ψ for any state of the system is shown in equation 2.21.

Equation 2.21 $$\psi = C_0\Phi_0 + \sum_{a,p} C_a^p \Phi_a^p + \sum_{\substack{a<b \\ p<q}} C_{ab}^{pq} \Phi_{ab}^{pq} + \sum_{\substack{a<b<c \\ p<q<r}} C_{abc}^{pqr} \Phi_{abc}^{pqr} + \ldots$$

where Cs are expansion coefficients, Φ is the spinorbital and $C_0\Phi_0$ is the HF level.

Adding the other configurations onto the zero order wavefunction $C_0\Phi_0$ allows the electrons to the higher virtual orbitals generating more movement space for the electrons and better separation between the electrons than in HF method. If a complete basis set is used, which is called full CI, any electronic wavefunction can be modelled exactly. However, using an infinite set of basis functions is computationally impossible to deal with. Therefore, a truncated basis set, which is named as limited CI, is employed at a computationally tractable level. CIS denotes a CI calculation that only contains Φ_0 and singly excited determinants with respect to Φ_0. CIS generates the same quality results as HF method because single excitation does not mix with the ground-state, at least double excitations (CISD) are needed to improve the HF predicted energies.

The important advantage of CI is that it is variational, and the disadvantage is the lack of size-consistency for the limited CI. Size-consistency means the energy of a system is proportional to the number of electrons n in the limit of $n \to \infty$. For medium to large molecules, only CIS is feasible in practice. CISD is only suitable to small to medium sized molecules. For the molecules investigated in this thesis, only CIS is feasible due to the available computing resources and computing cost.

2.3.3.1.1 Excited-state Configuration Interaction

CIS can be used to optimize a particular excited-state and then obtain the vertical absorption or emission energy. CIS provides satisfactory qualitative results, but it tends to overestimate excitation energies systematically due to the neglect of electron correlation as described above (CIS = HF quality), which leads to the electrons being closer than the real life. The transition energies calculated from CIS should be scaled by a factor of 0.72[52-54].

2.3.3.2 Møller-Plesset Many-body Perturbation Theory

Møller and Plesset took the zero-order Hamiltonian from the Fock operators of the HF method and applied a perturbation to the HF system in order to include the correlation energy of the ground-state. This method is known as Møller-Plesset perturbation theory (MPPT). MPPT overcomes the deficiency of limited CI, which is size-consistency.

The perturbation energy E of the system is given in Equation 2.22.

Equation 2.22 $$E = E_0 + \lambda E_0^{(1)} + \lambda^2 E_0^{(2)} + \ldots$$

where E_0 is the sum of the orbital energies, λ is a parameter describing the strength of the perturbation, $E_0^{(1)}$ is the energy of the first order correction, $E_0^{(2)}$ is the energy of the second order correction, and so on.

E_0 and $E_0^{(1)}$ together gives the ground-state HF energy. Therefore, $E_0^{(2)}$ is the first term to improve the HF description. MPPT theory is usually represented by MPn with n being the order of perturbation, therefore adding second-order energy correction to MPPT is designated MP2. Since the value of $E_0^{(2)}$ is always negative, The MP2 energy is smaller than the HF energy. Since MPPT is not variational as CI, the high order energy corrections may be positive that causes the overcorrecting of the energy. The existing of two-electron integrals increases the computational costs, as a result, this theory is usually applied to smaller systems.

2.3.4 Density Functional Theory

The *ab initio* methods discussed above have limitations. They need more time and more demanding computational effort to perform accurate calculations with using large basis sets on many-atom molecules. Density functional theory (DFT) attempts to solve these problems.

The basis of DFT is the use of electron probability density ρ to represent the energy of an electronic system, rather than a wavefunction. The ground-state energy and all the electronic properties of a molecule can be determined from the electron density. It is equivalent to say that electronic energy E is a functional of the electron density, designated E[ρ]. Kohn and Sham (KS)[55] showed the exact ground-state electronic energy E of an n-electron system, as given by Equation 2.23.

Equation 2.23

$$E[\rho] = -\frac{\hbar^2}{2m_e}\sum_{i=1}^{n}\int \psi_i^*(r_1)\nabla_1^2\psi_i(r_1)dr_1 - \sum_{I=1}^{N}\int \frac{Z_I e^2}{4\pi\varepsilon_0 r_{I1}}\rho(r_1)dr_1 + \frac{1}{2}\int \frac{\rho(r_1)\rho(r_2)e^2}{4\pi\varepsilon_0 r_{12}}dr_1 dr_2 + E_{XC}[\rho]$$

where the first term of the right hand side of the equation is the kinetic energy resulting from electron movement, the second term is the potential energy arising from the nucleus-electron attraction, the third term is the energy from electron-electron repulsion at r_1 and r_2, and the last term is the exchange-correlation term comprising the other electron-electron interactions. Spatial orbitals ψ_i (i=1, 2,…, n) are the Kohn-Sham (KS) orbitals. ρ(r) is the electron density at a location r.

The KS equations are solved in an iterative way that is similar to the SCF method. First the charge density ρ is deduced mostly from a superposition of atomic electron probability densities. Second, a functional form for electron-correlation energy depending on the electron density is chosen. In this work, a hybrid functional that is a linear combination of Beckes[56] three-parameter exchange functional and Lee, Yang and Parr[57] correlation functional (B3LYP) is used. Next, the exchange-correlation potential is computed. And then, KS orbitals are obtained by solving KS equations and a better charge density ρ is achieved. The process is repeated until the electron density converges to some specified tolerance. The electronic energy is then obtained by using Equation 2.23.

DFT method is similar in accuracy to MP2 method with using the same basis set, but with computational cost and efficiency closer to HF method. The energy of DFT

depends on electron density, whereas that of HF depends on wavefunctions. Therefore, DFT is less reliant on the size of the basis set than HF for obtain the accurate results, which in turn reduces the computational cost. The computational time for a DFT calculation is usually scaled as the third power of the number of basis functions.

2.3.4.1 Excited-state Time-dependent DFT Method

The time-dependent DFT (TDDFT) method is an extension to the DFT method; therefore the theoretical and computational foundations are similar between these two methods. The excited-state of a molecule can be explored by TDDFT of which the principle is that the electron density changes with varying external time-dependent potentials and has poles at the exact excitation energies of a system. The foundation of TDDFT is the Runge-Gross theorem[58] and Hohenberg-Kohn theorem[59].

2.3.5 Excited Electronic States Calculations

The analysis of simultaneous electronic and vibrational (vibronic) transitions is based on the Franck-Condon (FC) principle, such that the nuclear positions do not change during the transition and will readjust after the electrons adopt their final distribution. The optimal geometries of the ground-state and the excited-state may be different, which is shown in Figure 2.14. The vertical absorption transition is from equilibrium geometry of ground-state to the FC excited-state, which is the same geometry as the ground-state but projected onto the excited-state potential energy surface. The vertical emission transition is from the equilibrium geometry of excited-state to the excited-state geometry projected onto the ground-state potential energy surface.

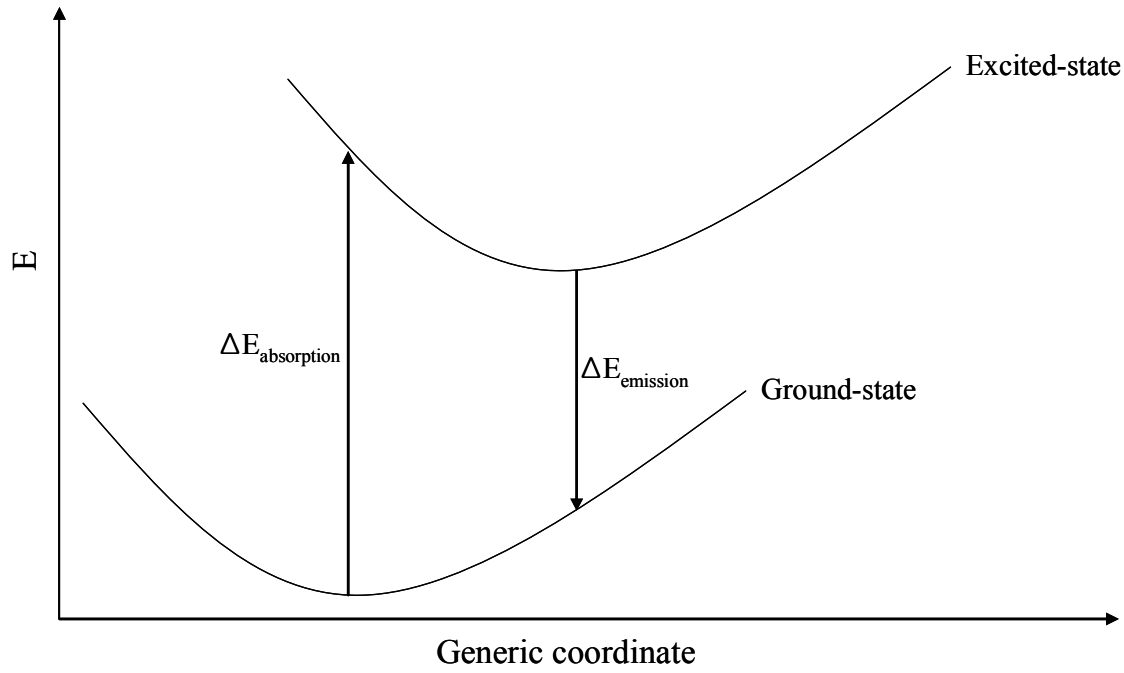


Figure 2.14 Diagrams of the vertical absorption and emission energies.

The two methods used in this work to perform excited-state calculations: CIS and TDDFT, which are previously discussed in Sections 2.3.3.1.1 and 2.3.4.1.

2.3.6 Basis Set Effects

So far, only the construction of the Hamiltonian has been discussed. However, how to build the other central part of the calculation, the wavefunction to describe the appropriate spin orbitals or KS orbitals, is still a question. Mathematically, any known functions can be linearly combined to form an unknown function, which could be used to describe to spin orbitals or KS orbitals. Each spin orbital or KS orbital is extended as a linear combination of known basis functions or in a basis set of series chosen functions.

A basis set is the mathematical description of the spin or KS orbitals (then combined to form the electronic wavefunction) within a system. To form the orbitals, a basis set typically assigns a group of basis functions to each atom within a molecule. These spin orbitals or KS orbitals are formed by a linear combination of a series of basis

functions. The accuracy of a calculation depends on the number and type of functions applied to the systems investigated.

An infinite set of basis functions should generate the optimal description of the wavefunctions. In practice, an infinite basis set is not feasible. The error resulting from the using a finite basis set is referred as basis set truncation error. Choosing a suitable number of basis functions for a system is a compromise between computational cost and accuracy, which means to keep the number of basis functions as small as possible and make sure the functions describe the systems correctly.

There are two main types of basis functions: one is Slater type orbitals (STO); the other one is Gaussian type orbitals (GTO).

STOs have some attractive features that with increasing the atomic radius they have the correct exponential decay and closely resemble hydrogenic atomic orbitals. However, in HF method, STOs have a significant limitation that the evaluation of the many two-electron integrals has no analytical solution available with three or more atoms.

GTOs are an alternative to STOs and makes the *ab initio* calculations feasible, as proposed by Boys in 1950[60]. For GTOs, the two different Gaussian functions at two different centres can be combined to form another Gaussian function with the centre between the two atoms. Therefore, the complex two-electron integrals with three or more different atomic centres can be reduced to integrals over two different centres. This makes the calculation computationally feasible. The disadvantage of GTOs is that the shape of s type orbitals ($i = j = k = 0$) is smooth, has zero slope at the nucleus ($r = 0$) and drops too quickly with increasing distance from the nucleus, whereas the real atomic orbitals have a cusp (discontinuous derivative) at the nucleus. Due to this problem, a larger basis set is needed to obtain a similar result to that achieved from STOs.

In order to overcome this disadvantage, contracted Gaussian type ortibal (CGTO), which is a linear combination of several primitive GTOs centred on the same atomic nucleus, are used in practice to represent the STO. A CGTO formed by three primitive GTOs are found to model the STO satisfactory. This basis set is termed as STO-3G, which is the simplest and minimal basis set. However, the results from the minimal basis set are not very accurate and not suitable for research purpose, thus, larger basis sets are needed to obtain more accurate results.

The first way to enlarge the basis set is to increase the number of basis functions for each atom. In the double zeta basis set (DZ), two basis functions are used for each atomic orbital. Similarly, in a triple zeta basis set (TZ), three basis functions are used to represent each orbital. DZ and TZ improve the accuracy significantly, as a consequence, both DZ and TZ result in a dramatically increase in computational cost compared with a minimal basis set. In order to reach a compromise to this problem, a split valence (SV) basis set is employed.

A SV basis set uses both the minimal basis set and DZ or TZ basis sets. It means that each of the valence atomic orbitals, which are most chemically interesting, is represented by two or three basis functions; and each inner shell atomic orbital, which is energetically essential but less chemically interesting, is represented by a single basis function (one contracted function). The restriction on the number of functions used to describe the inner shell atomic orbitals reduces the computational cost dramatically, and allows attention to be concentrated on the valence electrons. In a 3-21G basis set, ‘3’ indicates the inner shell atomic orbitals are described by a CGTO containing three primitive GTOs (PGTOs), which is the same as the STO-3G basis set. However, each valence orbital has two functions: ‘2’ means one function includes two PGTOs and ‘1’ means the other function includes one PGTO.

The atomic orbitals centred on atoms are distorted when chemical bonds form in molecules. Therefore, there can be insufficient mathematical flexibility for the wavefunction to describe the bonding in highly polar molecules. Additional flexibility can be accomplished by adding the polarization functions to a DZ basis set

to form a double zeta plus polarization (DZP) basis set. DZP takes into account of possible contributions from functions that represent orbitals with high values of angular momentum, beyond what is required for the ground-state, to the description of each atom. Therefore, DZP provides some angular flexibility to the basis set. For example, the distortion of p-orbital on the second row atoms of carbon can be corrected by adding d-type basis function.

Diffuse functions that enlarge the s - and p- type functions allow orbitals to occupy a larger region of space; therefore they are important for system where electrons are comparatively far from the nucleus, for example anionic systems with negative charges and molecules with many adjacent lone pairs of electrons. The diffuse functions are represented by the sign '+'.

In order to clearly illustrate the definition of a basis set, a basis set 6-311+G(d,p) is used as an example, the '6' indicates that each core orbital consists of six PGTOs; each valence orbital has three functions, hence, TZ basis functions; the '3' indicates that one function has 3 PGTOs; the two '1' indicate the other two functions consist of a single uncontracted PGTO each. The 'G' represents 'Gaussian' and means the type of the primitive function. The 'd' in brackets indicates that a group of polarization d-functions has been added to heavy atoms and 'p' in brackets means that a set of polarization p-functions has been added to hydrogen. The '+' indicates that diffuse s -, p - and d-functions have been added to heavy atoms.

A larger basis set puts fewer restrictions on the electrons location and generates more accurate orbitals. With increasing the basis set, the difference between HF energy and HF limit, which is obtained by using an infinite number of basis functions, decreases. The relationship of basis sets and computational methods, and the accuracy of these methods are illustrated in Figure 2.15.

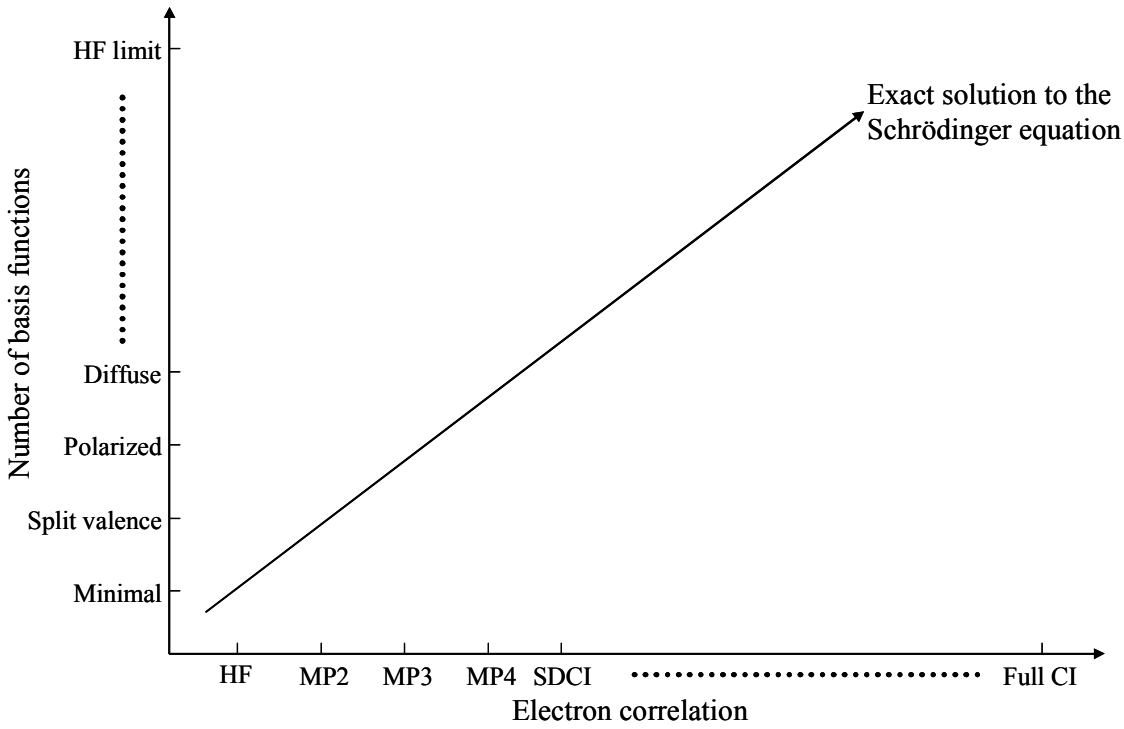


Figure 2.15 The relationship of basis sets and computational methods, and the accuracy of these methods.

2.3.7 Models for Solution Phase

So far, all the computational methods discussed above apply to gas phase molecules. However, many molecular systems of interest, especially in biochemistry, are in solution phase. A surrounding condensed phase can result in massive effects on the electronic structure of a molecule. As a consequence, any property relating to the electronic structure may have a different value between the solution and gas phase. The extent of the difference depends on the strength of the solute-solvent interactions. For example, the dipole moment of a molecule in solution is generally larger than that in the gas phase. The reason for this is that after dissolving the molecule in a solvent, the dipole of this molecule interacts with the electric field resulting from the dipole of the solvent, which helps to stabilize and enlarge the charge separation in the solute.

When a solute is dissolved in a solvent, the charge distributions of the solute and solvent interact with each other. There are two main models to modeling a solvent: one is continuous (implicit) model; the other one is explicit model. In continuous model, the solvent, which is considered as a homogenous medium, surrounds the solute molecule and the charge distribution of the solvent is represented by a continuous electric field named as reaction field (RF). In explicit solvation model, numbers of individual solvent molecules are added to surround a single solute molecule. The explicit model is more easily interpreted and closer to physical reality than the continuous model. However, there is a huge rise in computational complexity in explicit model. Because of the disadvantage of explicit model, in this work, the continuous model is employed to add solvent effect to a solute, of which the cavity is created from a series of overlapped atomic spheres centred on the atoms. The effect of solvent polarization is treated continuously on the surface of the cavity. The most broadly and frequently used method for the solute cavity is the surface area boundary element approach; first represented by Miertus, Scrocco and Tomasi in 1981[61]. They named their construction as the polarized continuum model (PCM).

Solvation is included in the Schrödinger equation by including an extra potential energy term due to the solvent RF and solving the complete equation self consistently in the standard HF way, which is named as a self-consistent reaction field (SCRF) calculation.

2.4 References

(1) Grinvald, A.; Steinberg, I. Z. *Anal. Biochem.* **1974**, *59*, 583-598.

(2) Badea, M. G.; Brand, L. *Method. Enzymol.* **1979**, *61*, 378-425.

(3) Ware, W. R.; Doemeny, L. J.; Nemzek, T. L. *J. Phys. Chem.* **1973**, *77*, 2038-2048.

(4) Beechem, J. M.; Knutson, J. R.; Ross, J. B. A.; Turner, B. W.; Brand, L. *Biochemistry* **1983**, *22*, 6054-6058.

(5) Beechem, J. M.; Ameloot, M.; Brand, L. *Chem. Phys. Lett.* **1985**, *120*, 466-472.

(6) Knutson, J. R.; Beechem, J. M.; Brand, L. *Chem. Phys. Lett.* **1983**, *102*, 501-507.

(7) Beechem, J. M. *Chem. Phys. Lipids* **1989**, *50*, 237-251.

(8) Beechem, J.; Gratton, E.; Ameloot, M.; Knutson, J.; Brand, L. *Topics in Fluorescence Spectroscopy*; Springer: 1991, *2*, 241-305.

(9) Beechem, J. *Method. Enzymol.* **1992**, *210*, 37-54.

(10) Watson, J. D.; Crick, F. H. C. *Cold Spring Harb. Sym.* **1953**, *18*, 123-131.

(11) Hunter, C. A. *BioEssays* **1996**, *18*, 157-162.

(12) Hunter, C. A. *Angew. Chem. Int. Edit.* **1993**, *32*, 1584-1586.

(13) Peon, J.; Zewail, A. H. *Chem. Phys. Lett.* **2001**, *348*, 255-262.

(14) Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem.* **1969**, *244*, 1228-1237.

(15) Eritja, R.; Kaplan, B. E.; Mhaskar, D.; Sowers, L. C.; Petruska, J.; Goodman, M. F. *Nucleic Acids Res.* **1986**, *14*, 5869-5884.

(16) Evans, K.; Xu, D.; Kim, Y.; Nordlund, T. M. *J. Fluoresc.* **1992**, *2*, 209-216.

(17) Rachofsky, E. L.; Osman, R.; Ross, J. B. A. *Biochemistry* **2001**, *40*, 946-956.

(18) Kawai, M.; Lee, M. J.; Evans, K. O.; Nordlund, T. M. *J. Fluoresc.* **2001**, *11*, 23-32.

(19) Stivers, J. T. *Nucleic Acids Res.* **1998**, *26*, 3837-3844.

(20) Rachofsky, E. L.; Seibert, E.; Stivers, J. T.; Osman, R.; Ross, J. B. A. *Biochemistry* **2001**, *40*, 957-967.

(21) Neely, R. K.; Jones, A. C. *J. Am. Chem. Soc.* **2006**, *128*, 15952-15953.

(22) Bonnist, E. Y. M.; Jones, A. C. *ChemPhysChem* **2008**, *9*, 1121-1129.

(23) Su, T.; Connolly, B. A.; Darlington, C.; Mallin, R.; Dryden, D. T. F. *Nucleic Acids Res.* **2004**, *32*, 2223-2230.

(24) Allan, B. W.; Reich, N. O. *Biochemistry* **1996**, *35*, 14757-14762.

(25) Reddy, Y. V. R.; Rao, D. N. *J. Mol. Biol.* **2000**, *298*, 597-610.

(26) Holz, B.; Weinhold, E.; Klimasauskas, S.; Serva, S. *Nucleic Acids Res.* **1998**, *26*, 1076-1083.

(27) Neely, R. K.; Tamulaitis, G.; Chen, K.; Kubala, M.; Siksnys, V.; Jones, A. C. *Nucleic Acids Res.* **2009**, *37*, 6859-6870.

(28) Neely, R. K.; Daujotyte, D.; Grazulis, S.; Magennis, S. W.; Dryden, D. T. F.; Klimasauskas, S.; Jones, A. C. *Nucleic Acids Res.* **2005**, *33*, 6953-6960.

(29) Lenz, T.; Bonnist, E. Y. M.; Pljevaljcic, G.; Neely, R. K.; Dryden, D. T. F.; Scheidig, A. J.; Jones, A. C.; Weinhold, E. *J. Am. Chem. Soc.* **2007**, *129*, 6240-6248.

(30) Hariharan, C.; Bloom, L. B.; Helquist, S. A.; Kool, E. T.; Reha-Krantz, L. J. *Biochemistry* **2006**, *45*, 2836-2844.

(31) Neely, R. K.; Magennis, S. W.; Dryden, D. T. F.; Jones, A. C. *J. Phys. Chem. B* **2004**, *108*, 17606-17610.

(32) Guest, C. R.; Hochstrasser, R. A.; Sowers, L. C.; Millar, D. P. *Biochemistry* **1991**, *30*, 3271-3279.

(33) Rist, M.; Wagenknecht, H.; Fiebig, T. *ChemPhysChem* **2002**, *3*, 704-707.

(34) Neely, R. K.; Magennis, S. W.; Parsons, S.; Jones, A. C. *ChemPhysChem* **2007**, *8*, 1095-1102.

(35) Kelley, S. O.; Barton, J. K. *Science* **1999**, *283*, 375-381.

(36) Wan, C.; Fiebig, T.; Schiemann, O.; Barton, J. K.; Zewail, A. H. *Proc. Nat. Acad. Sci.* **2000**, *97*, 14052-14055.

(37) O'Neill, M. A.; Barton, J. K. *Proc. Nat. Acad. Sci.* **2002**, *99*, 16543-16550.

(38) O'Neil, M. A.; Barton, J. K. *J. Am. Chem. Soc.* **2002**, *124*, 13053-13066.

(39) O'Neill, M. A.; Becker, H.; Wan, C.; Barton, J. K.; Zewail, A. H. *Angew. Chem. Int. Edit.* **2003**, *42*, 5896-5900.

(40) Steenken, S.; Jovanovic, S. *J. Am. Chem. Soc.* **1997**, *119*, 617-618.

(41) O'Neill, M. A.; Dohno, C.; Barton, J. K. *J. Am. Chem. Soc.* **2004**, *126*, 1316-1317.

(42) Nakatani, K.; Dohno, C.; Saito, I. *J. Am. Chem. Soc*. **2001**, *123*, 9681-9682.

(43) Shao, F.; Augustyn, K.; Barton, J. K. *J. Am. Chem. Soc.* **2005**, *127*, 17445-17452.

(44) Fiebig, T.; Wan, C.; Zewail, A. H. *ChemPhysChem* **2002**, *3*, 781-788.

(45) O'Neil, M. A.; Barton, J. K. *J. Am. Chem. Soc.* **2004**, *126*, 11471-11483.

(46) Bhattacharya, P. K.; Barton, J. K. *J. Am. Chem. Soc.* **2001**, *123*, 8649-8656.

(47) O'Neill, M. A.; Barton, J. K. *J. Am. Chem. Soc.* **2004**, *126*, 13234-13235.

(48) Neely, R. K., PhD thesis, The University of Edinburgh, **2005**.

(49) Nordlund, T. M.; Andersson, S.; Nilsson, L.; Rigler, R.; Graeslund, A.; McLaughlin, L. W. *Biochemistry* **1989**, *28*, 9095-9103.

(50) Hochstrasser, R. A.; Carver, T. E.; Sowers, L. C.; Millar, D. P. *Biochemistry* **1994**, *33*, 11971-11979.

(51) Bonnist, E. Y. M., PhD thesis, The Univeristy of Edinburgh, **2008**.

(52) Shukla, M. K.; Mishra, P. C. *Chem. Phys.* **1999**, *240*, 319-329.

(53) Kushwaha, P. S.; Kumar, A.; Mishra, P. C. *Spectrochim. Acta A* **2004**, *60*, 719-728.

(54) Yi, H.; Li, X.; Yang, S.; Duan, X. *Int. J. Quantum Chem.* **2003**, *94*, 23-35.

(55) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

(56) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(57) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(58) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997.

(59) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.

(60) Boys, S. F. *Proc. Roy. Soc. Lond. A Mat.* **1950**, *200*, 542-554.

(61) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117-129.

Bibliographies

Bevington, P. R.; Robinson, D. K. *Data reduction and error analysis for the physical sciences*; McGraw-Hill; 2nd edition: 1992.

Taylor, J. R. *An introduction to error analysis: the study of uncertainties in physical measurements*; University Science Books: 1982.

Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Springer; 3rd edition: 2006.

Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*; W.H. Freeman and Co; 6th edition: 2006.

Calladine, C. R.; Drew, H. R.; Luisi, B. F.; Travers, A. A. *Understanding DNA The Molecule and How it Works*; Elsevier Academic Press; 3rd edition: 2004.

Atkins, P. W.; de Paula, J. Friedman, R. S. *Quanta, Matter, and Change A molecular approach to physical chemistry*; Oxford University Press: 2009.

Atkins, P. W.; Friedman, R. S. *Molecular Quantum Mechanics*; Oxford University Press: 1997.

Foresman, J. B.; Frisch, A. E. *Exploring Chemistry with Electronic Structure Methods*; Gaussian Inc: 1996.

Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons, Ltd; 2nd edition: 2004.

Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to advanced electronic structure theory*; Courier Dover Publications: 1996.

Pilar, F. L. *Elementary quantum chemistry*; Courier Dover Publications; 2nd edition: 2001.

Chapter 3

Experimental and Computational Methods

3.1 Materials and Sample Preparation

2-aminopurine (2AP) (≥99%) was purchased from Sigma-Aldrich and 2-aminopurine riboside (2APr) from Carbosynth. They were used as received.

2AP-labelled DNA samples were provided by collaborators as indicated in the appropriate chapters. 2AP-containing dinucleotides were purchased from NEW ENGLAND BioLabs Inc. 2AP-containing thymine and 2AP-inosine were obtained from ATDBio Ltd. The fluorescence steady-state spectra of all solvents were measured first to insure that they did not contain fluorescent impurities. All solvents were spectrophotometric grade with purity no less than 99%.

The DNA duplexes were annealed before use. A solution of complementary oligonucleotides in 10 mM Hepes-NaOH (pH 7.5), 100 mM NaCl and 1 mM EDTA was heated in an Eppendorf tube, in a water bath, to 95 °C, held at this temperature for approximately 5 minutes, and then left to cool down slowly to room temperature.

4-selenium-thymidine-3'-phosphate and 6-selenium-2'-deoxyguanosine-3'-phosphate were gifts from Prof. Zhen Huang, Georgia State University.

3.2 Steady-state Fluorescence Measurements

Steady-state fluorescence spectra were recorded using a Fluoromax spectrofluorometer (Horiba Jobin Yvon) under the full control of DataMax

spectroscopy software. The FluoroMax is equipped with modified Czerny-Turner spectrometers in both the excitation and emission positions. In both excitation and emission spectrometers a grating disperses the light from 200 to 900 nm and has a groove density of 1200 grooves/mm. By adjusting the widths of slits of each spectrometer, the intensity and wavelength spread (bandpass) of the light can be controlled. In this work, the slits for both excitation and emission spectrometers were kept the same and corresponded to a 5 nm bandpass unless specified.

Before the light enters the sample compartment, a beam splitter directs 8% of the light to a reference photodiode and the remaining light continues to the sample to excite fluorescence. The reference photodiode is used to monitor the excitation source (xenon lamp), correct temporal fluctuations and the wavelength-dependence of the lamp intensity, and compensate for the wavelength response characteristics of the excitation spectrometer. By ratioing the fluorescence signal to the reference signal, correction of measured signal intensities were made for variations in excitation light intensity.

Light exiting the emission spectrometer is monitored by a photomultiplier detector operating in the photon-counting mode, and data are acquired in counts per second. The photomultiplier detector saturates at 4×10^6 counts per second. When the fluorescence intensity exceeded this value, the bandpass was reduced until the maximum intensity fell below this value in order to obtain a linear response of the photomultiplier detector.

The wavelength increment determines the number of data points collected. An increment of 1 nm was used in this work unless specified. The integration time determines the length of time for which data is collected for each point. Selection of an appropriate integration time is based on the signal intensity of the fluorescence and the signal to noise ratio. Integration times between 0.1 and 1 second per wavelength interval were used in this work, the longer integration time used for the lower signal intensities in order to obtain sufficiently high signal to noise ratio.

Samples measured at room temperature were contained in a fused silica fluorescence micro cuvette (Starna) with capacity 60 µl, a 2 x 2 mm excitation window and a 2 x 10 mm emission window. The arrangements for samples measured at 77 K are described in Section 3.4.

3.3 Time-resolved Fluorescence Measurements

Fluorescence lifetimes were measured using the time-correlated single photon counting technique (TCSPC), in the Collaborative Optical Spectroscopy Micromanipulation and Imaging Centre (COSMIC) at the University of Edinburgh. A schematic of the experimental setup is shown in Figure 3.1.

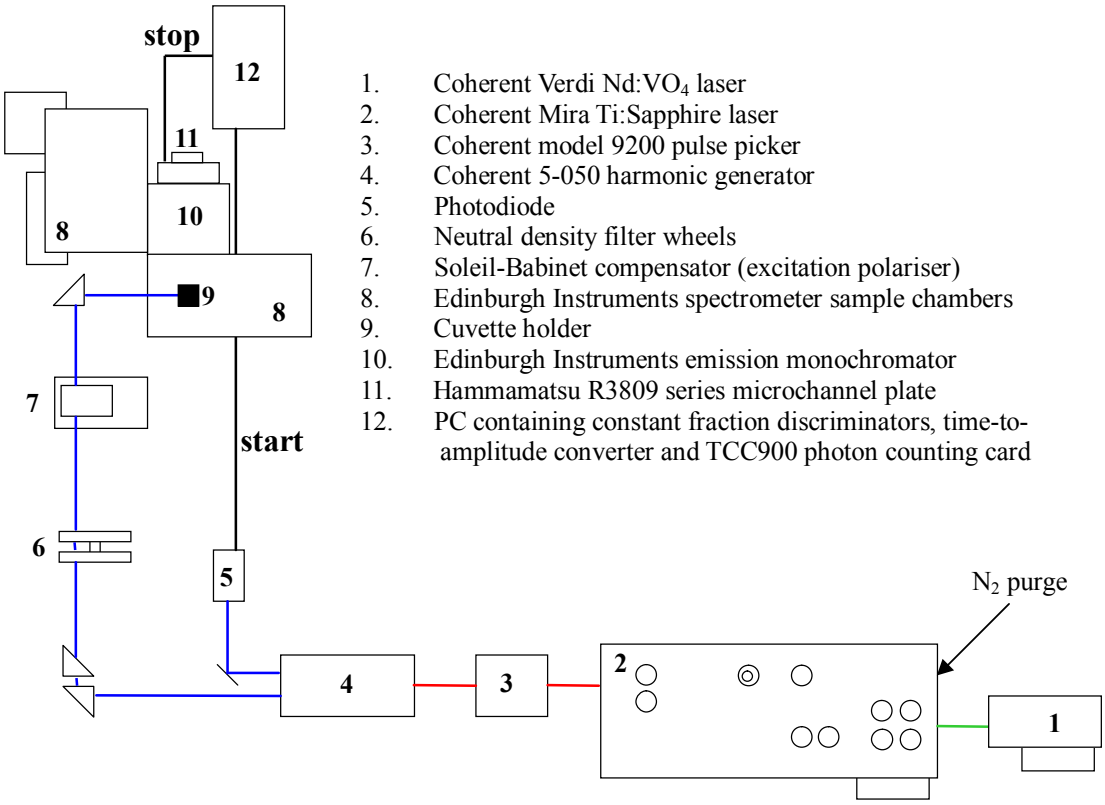


Figure 3.1 Schematic of equipments used in experiments. Reproduced from reference 1.

A Coherent Verdi V-10 Neodymium: Vanadate (Nd: YVO_4) continuous wave solid state laser with a power of 10 W was used to pump a mode-locked Ti:Sapphire laser (Coherent Mira) with tuning range from 700 to 1000 nm. The Ti:Sapphire laser

produces pulses at a repetition rate of 76 MHz and with a width of ~200 fs. A pulse picker (Coherent 9200) reduces the repetition rate to 4.75 MHz, which means a pulse is emitted every 210 ns. The light output was directed into a harmonic generator (Coherent 5-050) where the wavelength was reduced to half or one-third in order to obtain the desired excitation wavelength. Part of the laser output from the harmonic generator was diverted to a fast photodiode, which converted the laser pulse to an electronic pulse. This pulse triggers START signal for the time-to-amplitude converter (TAC). The amplitude of the photodiode output pulse was set to 0.7~0.8 V. The intensity of the excitation beam can be reduced with neutral density filters.

Vertically polarised laser light was obtained by passage through a Soleil Babinet Compensator, which modifies the orientation of linear polarised laser light. Depending on the laser wavelength, the Soleil Babinet Compensator needs to be tuned to the appropriate position. Prior to performing an experiment, to ensure correct position of the Soleil Babinet Compensator, a resolution value, which determines the travel distance between 'no change to the laser orientation' and '90 degree change to the laser orientation' was set in the controlling software (Edinburgh Instruments (EI) F900). Equation 3.1 shows the relationship between resolution value (R) and wavelength λ.

Equation 3.1 $$R = \frac{41.62}{\lambda}$$

The vertically polarised light was directed into the EI spectrometer chamber and focussed in the centre of the cuvette holder. Fluorescence steady-state spectra were measured using an EI photon counting photomultiplier tube. Fluorescence decay curves were collected using a Hamamstu R3809-50 series microchannel plate photomultiplier (MCP). To avoid any anisotropic effects on the results, fluorescence emission was directed through a polariser set at the magic angle 54.7 ° before reaching the Hamamstu MCP. The fluorescence emission wavelength was set by using an EI M300 monochromator with slits set corresponding to 10-15 nm bandpass, depending on the emission intensity, the lower intensity the wider slits. To insure

complete elimination of scattered laser light from the sample, a bandpass filter 377 nm (352-402 nm bandpass) or longpass filter (Schott) were placed between the sample and monochromator. The STOP signal for TAC was generated by detection of a fluorescence photon by the MCP detector.

The system alignment was optimised by scattering the excitation laser beam from a dilute Ludox solution held in a 5 ml capacity fused silica transparent fluorescence cuvette in the spectrometer chamber. This Ludox solution was also used for measuring the instrument response function (IRF). The typical IRF of the TCSPC in this work is 80 ps full width at half maximum intensity (FWHM). The optimisation of the system was further checked by measuring the known single-exponential fluorescence lifetime of 2APr. The stop count rate of samples was kept below 30,000 Hz in order to avoid damage to the MCP.

Fluorescence decay curves were measured with EI F900 TCSPC electronics. All the parameters for the experiments were set through this software. Before measuring the fluorescence decay curves, the steady-state spectra of the samples were recorded to check the state of the samples and determine the emission wavelengths for the decay curves. The three emission wavelengths are normally chosen from the peak of the emission spectrum and moving 10-20 nm from each side. The decay data were accumulated to 10,000 or 20,000 counts in the maximum channel. Data were resolved into 4096 channels over a 50 ns range, corresponding to 12 ps per channel.

Individual decay curves were fitted using EI F900 software, whereas the global analysis of three decay curves was performed using EI FAST software. The fitted range was from the peak of IRF to the channel where the photon counts fell to approximately 10 (~10 x background). The decay background value used in the fitting was fixed and estimated from the average of photon counts for all the channels before the rising edge of the decay. Decay curves were fitted to a single exponential initially. If the quality of the fit by examining the residuals of the fit and the magnitude of the reduced χ^2 was not satisfactory, the decay curves were fitted to a bi-exponential. This procedure was repeated until an acceptable fit was reached. A

maximum of four lifetimes can be fitted to decay curves using F900 and FAST softwares.

3.4 Measurements at 77 K

In order to investigate the role of base dynamics in populating the conformational states and the nature of heterogeneous decay of 2AP excited states in dinucleotides, the fluorescence spectra and decays of 2AP-labelled dinucleotides were examined at 77 K.

2AP-labelled dinucleotides were dissolved in 10 M aqueous LiCl in order to form a stable and transparent glass at 77 K. 0.5 ml of the dinucleotide/ LiCl solution was then transferred into a NMR-style fused silica tube (made by the in-house glass blowing workshop) with a diameter of 5 mm. The tube was inserted very slowly into liquid nitrogen contained in a Horiba Jobin Yvon fused silica dewar, illustrated in Figure 3.2.

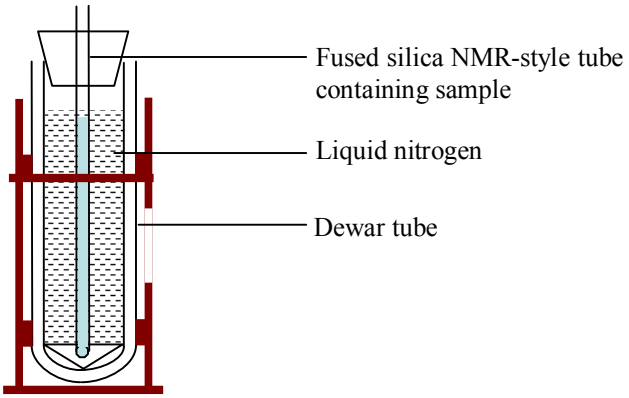


Figure 3.2 Arrangement for fluorescence measurements at 77 K. Reproduced from reference 1.

For the lifetime measurements at 77 K, the IRF was measured with Ludox in the sample tube, but no liquid nitrogen in the dewar. During the experiments, the level of liquid nitrogen was maintained above the level of the sample in the tube. Also, condensation on the dewar must be avoided to obtain good quality data. When the humidity was very high, a beaker of liquid nitrogen was placed in the chamber, to condense out water vapour.

3.5 Computational Methods

Computational calculations were used to study the molecular structures and spectroscopic properties of DNA natural bases and selenium-containing base analogues. The calculations included optimised geometry, vertical excitation energies, and wavelength and oscillator strength of excitation and emission spectra. All calculations were carried out using the Gaussian03[2] software package within the Edinburgh and St Andrews Research School of Chemistry (EaStCHEM) Research Computing Facility. ArgustLab, Gabedit, gOpenMol and ChemDraw software packages were used to visualise the computational results.

Ground state geometries were optimised using RHF, B3LYP and MP2 theories with 6-311+G(d,p) basis set separately (detailed discussion in Section 2.3). The reason for using different methods was to compare the accuracy of results among different methods.

Geometry optimisations do not necessarily only locate minima on the potential energy surface but may also locate transition state structures that correspond to a saddle point with a maximum in one coordinate and a minimum in the other coordinates. The first derivative of the energy is zero at both minima and saddle points. In order to distinguish between the minima and transition state structures of the molecule, a frequency calculation was performed on the optimised structure. Vibrational frequencies are the second derivatives of the energy corresponding to the nuclear positions. The absence of No negative (imaginary) frequencies indicates a minimum in the potential energy hyper-surface, whereas one or more imaginary frequencies indicates a saddle point. In order to find a minimum, using a saddle point as a starting point, the Cartesian co-ordinates of the first imaginary frequency were added to the Cartesian co-ordinates of the saddle point, and then the new co-ordinates were used to rerun the geometry optimisation and frequency calculations of which the step was repeated several times until no imaginary frequency exists.

Absorption spectra were calculated using TD-DFT with Becke's three-parameter exchange functional combined with Lee, Yang and Parr's correlation functional (B3LYP) and CIS with 6-311+G(d,p) basis set. The former was based on the ground state B3LYP optimised geometry, and the latter was based on the ground state MP2 optimised geometry. Fluorescence spectra were simulated using the same methods as above, followed by optimising the excited state molecular geometry using CIS (also see details in Section 2.3).

The effect of solvation (solvent is water) was simulated using a self-consistent reaction field polarised continuum model (SCRF-PCM) theory. For the optimised ground state geometry in water, the MP2 method with 6-311+G(d,p) basis set was used, based on the ground state MP2 optimised geometry in the gas phase. The absorption spectra in water were calculated using CIS with 6-311+G(d,p) basis set, also based on the ground state MP2 optimised geometry in the gas phase. The fluorescence spectra in water were calculated using TD-DFT (B3LYP) and CIS with 6-311+G(d,p) basis set, based on the optimised excited state geometry in solution using SCRF-PCM/CIS. All the optimised geometries were verified by frequency calculations.

The default convergence criteria for all calculations were: Maximum Force = 0.000450, RMS Force = 0.000300, Maximum Displacement = 0.001800 and RMS Displacement = 0.001200.

3.6 References

(1) Bonnist, E.Y.M., PhD thesis, The University of Edinburgh, **2008**.

(2) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr. T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H.

P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision E.01*, Gaussian Inc., Wallingford CT, **2004**.

Chapter 4

2-Aminopurine-labelled DNA Single Strands

4.1 Introduction

In recent years, 2AP has been widely used as a fluorescent probe to study DNA conformational changes and bases interactions. However, most of the studies have focused on the DNA helix. The conformational behaviour and interactions between bases in DNA single strands have been far less investigated. In order to compare the conformations, π-stacking interactions and base mobility in DNA double and single strands, and investigate the importance of hydrogen bonding on maintaining the DNA conformation, the fluorescence properties of six 2-aminopurine (2AP) - labelled single strands have been studied by steady-state and time-resolved fluorescence measurements, as reported in this chapter. The sequences of the single strands are listed in Table 4.1.

Single strand	Base sequence
APA	5'-TGC A**P**A ATT CGA GGT CGA CGG-3'
APT	5'-TGC AGA **P**TT CGA GGT CGA CGG-3'
CPA	5'-CCG TC**P** ACC TCG AAT TCT GCA-3'
GPA	5'-TGC AG**P** ATT CGA GGT CGA CGG-3'
GPT	5'-TGC AGA ATT CGA G**P**T CGA CGG-3'
GPC	5'-CCG TCG **P**CC TCG AAT TCT GCA-3'

Table 4.1 Base sequences of the DNA single strands, which are designated APB where P stands for 2AP, A and B are the neighbours of 2AP at 5' to 3' sides, respectively.

Bases paring between complementary bases and π-stacking interactions between adjacent base-pairs are the two interactions that contribute to the stability of the DNA duplex structure[1,2]. DNA stability plays an essential role in the function of DNA in the cell and in biotechnological applications, such as DNA sequencing and

polymerase chain reaction. The relative contributions of base paring and π-stacking interactions to the overall duplex stability was quantitatively estimated by Protozanova *et al.*[3]. The major hypothesis underlying their study was that nicked DNA (a single stranded break in the DNA backbone) exists, in solution or gel, in an equilibrium geometry between the closed and open states, governed by the stacking-unstacking free energy difference, ΔG_{KL}^{ST}, of neighbouring base-pairs flanking the nick, shown in Figure 4.1. The stacking between the two base-pairs flanking the nick is maintained in the closed state, very similar to the conformation of the intact DNA duplex. The stacking between the two base-pairs is lost in the open state, introducing a kink in the DNA duplex, which affects its electrophoretic migration. The free energy difference for the closed-open equilibrium at the site of the nick, for all 16 nicked dinucleotide stacks in 300 bp-long DNA fragments, was determined by using urea-enhanced polyacrylamide gel electrophoresis (PAGE) and by extrapolating the stacking free energy values to zero urea concentration[3]. The contribution of base-pairing and π-stacking interactions to the overall duplex stability had been estimated from these data and from a well-known dependence of DNA melting temperature on guanine (G) · cytosine (C) content. The results showed that the DNA double helix was stabilized by π-stacking interactions between adjacent base-pairs rather than base-pairing between complementary bases. This was attributed to the easy compensation of the lost of hydrogen bonding between the base-pairs by forming the hydrogen bonds between the separated bases and water molecules, whereas the π-stacking interactions between adjacent base-pairs were very hard to compensate by other interactions.

Figure 4.1 Closed and open states of the DNA fragment at the nick site described by the energy difference, ΔG_{KL}^{ST} . Reproduced from reference 3.

Non-covalent π-stacking interactions between adjacent bases and base-pairs in DNA play an important role in determining DNA properties and sequence-dependent structures[4]. The non-covalent π-stacking interactions consist of van der Waals (vdW) and electrostatic interactions, as described in Section 2.2.1 of Chapter 2. The electrostatic interactions are proposed not only to be dominant in controlling the extent of the π-stacking interactions for aromatic molecules[5-8], but also to play a key role in determining the orientation of π-stacking interactions[9]. The results from modified neglect of differential overlap (MNDO) calculations show that an electrostatic complementarity between partial charges on the two aromatic rings results a strong face-to-face stacking[9]. A weak edge-to-face stacking is favoured in the absence of the superimposed electronic charge distributions on the two planes of the aromatic molecules[9]. Besides that, the edges of aromatic rings cooperate better with groups containing negative charges than those having positive charges[7].

The stacking interactions as a function of key DNA structural parameters, twist and rise, for ten DNA Watson-Crick base-pairs, were investigated using density functional theory (DFT) computational calculations with van der Waals density functional (vdW-DF) for the exchange-correlation energy[10]. These calculations did not include the sugar-phosphate backbone or any solvent medium and also excluded roll, shift, slide, buckle and propeller parameters. The results showed that stacking interactions between nucleic acid base-pairs were crucial for defining the twist angles

(twist) and base-pair separation distance (rise). The twist and rise parameters are the consequence of the interaction between short-range Pauli repulsions and long-range vdW and electrostatic interactions[10]. The stacking energies (13.5 to 18.2 kcal mol^{-1}) are comparable to hydrogen-bonding energies between nucleobases (15.0 kcal mol^{-1} for adenine (A) : thymine (T)[11] and 27.5 kcal mol^{-1} for G:C[11]), suggesting the strength of stacking interaction is similar to that of hydrogen bonding interactions between nucleobase-pairs[10]. The base-pair stacking stability was found to change by 8-13 kcal mol^{-1} with varying rise parameter, and by 2-5 kcal mol^{-1} with change in twist parameter[10]. Therefore, base-pair stacking stability seems to be controlled by the rise parameter[10]. The results also implied that T offered additional stability to DNA due to the methyl group of T having stacking interactions with neighbouring base-pair steps. Stacking between AT:AT base-pairs increases the total stacking energy by up to ~15% , which in turn reduces twist angles substantially[10].

The fluorescence lifetimes of 2AP in single strands have been little studied until now. To the author's knowledge, the only previous study was by Ramreddy *et al.*[12] who investigated varying length single strands in the range of thirty to seventy repeating A bases with 2AP instead of A at various position (1, 2, 4, 15, 28 and 30) from the 5' end. The reason for using repeats of A rather than the other three natural bases was to make the neighbours on both sides of 2AP identical in different single strands and focus on the effects of base motion and stacking interaction at different positions in the sequence rather than the identity of the neighbouring bases of 2AP. Fluorescence intensity decays of 2AP-labelled DNA single strands with excitation wavelength 310 nm showed there were three fluorescence lifetimes, which were 0.5-0.7 ns, 1.6-2.4 ns and 4.2-6.3 ns. (The full width at half maximum (FWHM) of the instrument response function (IRF) was ~40 ps[12].) These lifetimes were attributed to 2AP in three different conformations of the single strands. The second lifetime accounted for more than 50% of the 2AP population and was dominant, whereas the longest lifetime had the lowest population, less than 20%. All three lifetimes, and the average lifetime, increased as 2AP was moved from the end to the middle of the strand. The quenching of 2AP fluorescence was attributed to stacking interactions between 2AP and its near neighbours and non-radiative processes induced by internal motion of 2AP within the

strand. The decrease of the mean fluorescence lifetime of 2AP at the end of the strand was attributed to the higher level of internal motion resulting in enhanced non-radiative decay. However, this could not explain the fluorescence behaviour of 2AP at the second position from 3' end which resembled that in the middle and differed from that at the second position from 5' end. The fluorescence excitation spectra of 2AP in the single strands had two peaks: one at 260 nm, arising from the energy transfer from neighbouring bases to 2AP, and one at ~320 nm, due to direct excitation of 2AP. The intensity of the 260 nm excitation band of 2AP was highest when 2AP was in the middle of the strand, indicating that base stacking interaction in the middle was stronger than that at the end of single stands.

Hardman *et al.*[13] performed CIS/3-21+G(d,p) computational calculations on double and single strand trinucleotide models (X2APX)·(YTY), where X and Y are the natural bases, in order to study the combined effect of base stacking and hydrogen bonding on the fluorescence transitions of 2AP. The sugar and phosphate parts were omitted from the calculation to keep the size of the molecule feasible for computational calculation. When only the base stacking effect was considered, it was found that the oscillator strength of fluorescence transition of 2AP in double and single strand trinucleotides decreased compared with that of 2AP alone. Hydrogen bonding alone caused an increase in the oscillator strength of the fluorescence transition in double and single strands. However, the combination of base stacking and hydrogen bonding led to a decrease in the oscillator strength compared with that of free 2AP[13]. The fluorescence emission peaks of 2AP-labelled double strand trinucleotides were predicted to be shifted to longer wavelength by 6-12 nm relative to free 2AP. The decrease of oscillator strength results in a decrease in radiative rate constant and hence reduced fluorescence quantum yield for the 2AP-labelled DNA double and single strand trinucleotides compared with that of 2AP alone. It was also suggested that non-radiative decay rates of 2AP-labelled DNA double and single strands might increase because of the orbital delocalization across 2AP and other bases, which would cause the fluorescence quantum yield to decrease further.

A previous study of long wavelength (450 nm) emission of 2AP in DNA (as described in Section 2.2.2.1 of Chapter 2) showed that this occurs in single strands, as well as in duplex DNA[14]. Long wavelength emission was detected in single strands containing 2AP stacked with each of the four natural ones. The long wavelength fluorescence decay of 2AP in duplex DNA was found to be multi-exponential, with four lifetimes, as described in Section 2.2.2.3 of Chapter 2, but the decay were not recorded for single strands. The latter have been studied in the present work. Although the mechanism, driving force, efficiency and other features of electron transfer (ET) between the natural bases and excited 2AP in duplex DNA have been studied thoroughly and are quite well-understand, as described in Section 2.2.2.2 of Chapter 2, there has been little study of ET in single strands. In this chapter, the fluorescence spectra and lifetimes of 2AP-labelled DNA single strands are presented and compared with those of duplexes. The role of ET in quenching 2AP fluorescence in single strands is elucidated.

4.2 Materials and Methods

2AP-labelled single strands were gifts from Prof. Geoff Kneale, the University of Portsmouth. The sequences of the six 2AP-labelled single strands are shown in Table 4.1. These single strands were buffered in 100 mM NaCl, 10 mM Tris-Cl, 1 mM EDTA at pH 8.0. The concentration of GPC was 16.7 μM; the concentration of the other five single strands was 20 μM.

Steady-state fluorescence was measured using a Fluoromax spectrofluorometer, as described in Chapter 3. The excitation spectra were recorded with emission wavelengths from 360 nm to 500 nm, at 10 nm intervals, whereas the emission spectra were measured with excitation wavelength from 290 nm to 400 nm, at 10 nm intervals. Spectral intensity was calculated by integrating over a 20 nm wide band centred on the wavelength of maximum emission intensity. Final values were converted to an equivalent concentration of 20 μM 2AP.

Time-resolved fluorescence was recorded using TCSPC as described in Chapter 3. The fluorescence lifetimes were measured at two excitation wavelength 309 nm and 360 nm. The 309 nm excitation radiation was produced by frequency tripling the 927 nm output of the Ti: Sapphire laser. The fluorescence decays were collected at three emission wavelengths: 370 nm, 385 nm and 400 nm (except for APA those of which were recorded at 370 nm, 390 nm and 400 nm). The 360 nm excitation radiation was produced by frequency doubling the 720 nm output of the Ti: Sapphire laser. The fluorescence decays were measured at 450 nm and 470 nm. The slits of monochromator were set to 10 nm bandpass.

The relative quantum yield (Φ_{rel}) of the sample with respect to free 2AP riboside (2APr) in the same solvent was calculated from the fluorescence decay parameters according to Equation 4.1. The quantum yield quoted is the average of the values at the three different emission wavelengths measured.

Equation 4.1 $$\Phi_{rel} = \frac{\langle \tau \rangle}{\tau_{2A\,\mathrm{Pr}}}$$

4.3 Results

4.3.1 Steady-state Fluorescence

The emission spectra of one of the single strands, GPT, following excitation at 310 nm and 360 nm are shown in Figure 4.2. Excitation at 310 nm generates an emission spectrum with maximum intensity at about 370 nm. Excitation at 360 nm produces a long-wavelength emission with maximum intensity at ~450 nm, but this emission is much weaker (~0.02) than the 370 nm emission spectrum.

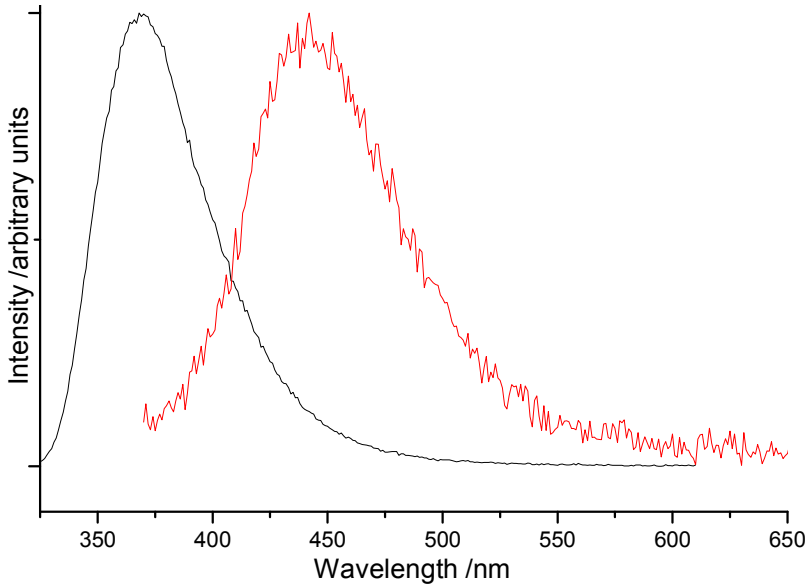


Figure 4.2 The emission spectra of GPT single strand, excited at 310 nm (black line) and 360 nm (red line, intensity multiplied by a factor of 50).

The excitation spectra of GPT with emission wavelengths 370 nm and 450 nm are illustrated in Figure 4.3. The excitation spectrum has a peak at 310 nm with emission wavelength 370 nm. With emission wavelength 450 nm, the excitation spectrum has a new shoulder at 360 nm, indicating the long wavelength emission is generated from a distinct ground state species. The persistence of the 310 nm peak in the excitation spectrum of the 450 nm emission has been discussed previously[14].

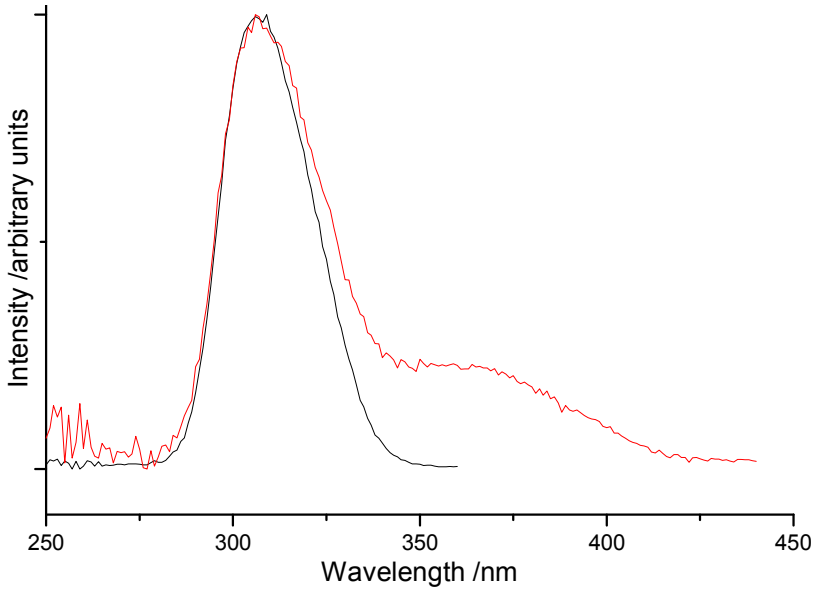


Figure 4.3 Excitation spectra of GPT with emission wavelengths 370 nm (black line) and 450 nm (red line, intensity multiplied by a factor of 50).

The fluorescence excitation and emission spectra of the other five 2AP-labelled DNA single strands are very similar to those of GPT and are not shown here. The 450 nm emission spectral properties are consistent with those measured previously by Bonnist *et al.*[14]. The present observations confirm the previous conclusions that that the heterodimer that gives rise to the long wavelength emission is formed through π-stacking interaction between 2AP and one of its adjacent bases, and occurs with any of the natural bases.

4.3.2 Fluorescence Lifetimes

4.3.2.1 370 nm Emission

Fluorescence decay parameters at excitation wavelength 309 nm are represented in Table 4.2. Amplitude parameters (A) are for emission wavelength 385 nm, except those for APA which are for 390 nm. The full set of the decay parameters for the six single strands at all three emission wavelengths given in Tables A1 to A6 in Appendix A.

Single strand	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	$\langle\tau\rangle$/ ns
APA	-	0.68	2.4	9.6	-	0.32	0.53	0.15	2.93
APT	0.11	0.52	2.1	7.2	0.42	0.39	0.13	0.06	0.95
CPA	0.12	0.55	1.7	5.5	0.41	0.35	0.20	0.04	0.80
GPA	0.06	0.59	2.4	6.9	0.50	0.19	0.22	0.09	1.29
GPT	0.05	0.53	2.7	7.5	0.69	0.11	0.10	0.10	1.11
GPC	0.08	0.59	2.4	6.9	0.45	0.16	0.31	0.08	1.43

Table 4.2 The global fluorescence decay parameters for single strands excited at 309 nm. A factors are for 385 nm emission, except those for APA which are for 390 nm. Average lifetimes are also shown. Experimental uncertainties (estimated by dividing standard deviation of values by the corresponding values): τ_1 (≤ 10%), τ_2 (≤ 5%), τ_3 (≤ 3%), τ_4 (≤ 1%), A_1 (≤ 2%), A_2 (≤ 3%), and A_3 and A_4 (≤ 4%).

It can be seen that fluorescence decay of 2AP in single strands is multi-exponential characterised (with the exception of APA) by four discrete components. This indicates the existence of different conformations, as observed in previous studies of 2AP-labelled duplexes (as described in Section 2.2.2.3 of Chapter 2). Except for APA, the fluorescence decays of are dominated (> 40%) by a short component of less than 120 ps, whereas in APA this shortest lifetime component is missing. Less than 10% of the 2AP molecules in the other five single strands have fluorescence lifetimes about 7 ns, whereas in APA the longest lifetime component of 2AP is ~10 ns with a slightly greater population than in the other five. The two intermediate lifetimes have similar values, ~0.6 ns and ~2.4 ns, for all the six single strands. However, the population of these two components varies considerably; for APA the 2.4 ns component is dominant. There is almost no dependence of the amplitudes of lifetimes on different emission wavelength (see Tables A1 to A6 in Appendix A), indicating that the decay components have very similar emission spectra.

The average fluorescence lifetimes of 2AP in DNA single strands are in the range of 0.8 to 3 ns, which are significantly smaller than that of free 2APr in aqueous solution 10.6 ns[15]. This indicates the fluorescence quantum yields of 2AP-labelled single strands are much reduced compared with that of free 2APr.

4.3.2.2 450 nm Emission

Fluorescence decay parameters for single strands excited at 360 nm are shown in Table 4.3. Amplitudes in Table 4.3 are for emission wavelength 450 nm. The full set of decay parameters are presented in Tables A7 to A11 in Appendix A.

Single strand	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	$<\tau>$/ ns
APA	0.12	0.59	3.6	14.1	0.23	0.57	0.11	0.09	2.03
CPA	0.18	1.20	4.5	14.2	0.32	0.26	0.25	0.17	3.91
GPA	0.15	1.10	4.3	15.1	0.39	0.24	0.14	0.23	4.40
GPT	0.22	2.20	7.2	16.9	0.25	0.19	0.46	0.10	5.48
GPC	0.14	1.40	5.7	15.0	0.31	0.22	0.24	0.23	5.17

Table 4.3 The global fluorescence decay parameters for single strands which are excited at 360 nm. A factors are for 450 nm emission. Average lifetimes are also shown. Experimental uncertainties (estimated by dividing standard deviation of values by the corresponding values): τ_1 ($\leq$ 12%), τ_2 ($\leq$ 6%), τ_3 and τ_4 ($\leq$ 3%), A_1 ($\leq$ 5%), A_2 ($\leq$ 4%), A_3 ($\leq$ 3%), and A_4 ($\leq$ 12%).

All the single strands have four discrete fluorescence lifetimes when they are excited at 360 nm, indicating the 2AP dimer experiences a heterogeneous environment. The magnitudes of all the four lifetimes are generally significantly greater than (about twice) those for 370 nm emission, indicating slower quenching rates, and the shortest component does not show the sequence-dependence seen in the 370 nm case. The amplitude of the shortest component is substantially less than for 370 nm emission and consequently the dimer population is more evenly distributed amongst the four conformational states. These observations are similar to those reported previously for duplex DNA[16], where is suggested that the geometrical constraint required to maintain the stacked dimer structure restricts the ability of the dimer to achieve highly overlapped and stacked conformations with its neighbouring bases.

The average lifetime shows significant sequence-dependence. The value of the average lifetime is mainly determined by the combined amplitudes of components 3 and 4, which are attributed to poorly stacked or unstacked conformations.

4.4 Discussion

4.4.1 Effect of Sequence Context

4.4.1.1 370 nm Emission

There is strong evidence that the dominant mechanism for quenching of 2AP in duplex DNA is ET from G. It will now be considered whether this holds true in single strands. Table 4.4 shows the proximity of 2AP to the nearest G in the 5' and 3' directions.

Single strand	Nearest 5' G/ bases	Nearest 3' G/ bases
APA	3	6
APT	2	4
CPA	3	6
GPA	1	5
GPT	1	3
GPC	1	5

Table 4.4 The proximity of 2AP to the nearest G for each single strand investigated. The number means the n^{th} base is G compared with central 2AP, for example '1' indicates G is the neighbouring base of 2AP.

GPA, GPT and GPC all have a very short-lived component, with lifetime less than 80 ps, which is consistent with rapid ET quenching from the flanking 5'G in each case.

Wan *et al.*[17] have proposed that efficient ET from G to 2AP* can occur over up to 3 intervening (bridging) bases if these are adenines, but ET is inhibited by intervening C or T. For the APT single strand, the shortest lifetime component of 110 ps would be consistent with ET from the 5'G across the bridging A. However, measurements on 2AP dinucleotides reported in Chapter 5 of this thesis show that 2AP can be rapidly quenched by charge transfer with 3'T, which could therefore also account for this short component in this single strand.

In the CPA single strand the 120 ps component can also be accounted for by charge transfer quenching by C, by comparison the 2AP-C dinucleotide reported in Chapter 5.

An interesting abnormal single strand is APA, which lacks the shortest lifetime component, suggesting that fast ET completely disappears. This is similar to the behaviour seen by Ramreddy *et al.*[12] for 2AP in poly-A single strands. For a 30-A strand with 2AP at the fourth position from the 5' side they report three fluorescence lifetimes 0.56 ns (0.25), 2.3 ns (0.63) and 5.8 ns (0.12), with the populations shown in the brackets. In APA, 2AP is at the fifth position from 5' side in a 21 different bases single strand. The similarity of the decay components indicates that in APA the shortest component (0.68 ns) can be attributed to quenching of 2AP* by the neighbouring adenines with no significant contribution from G. This is further supported by the observation of analogous quenching in the 2AP-A dinucleotide, as reported in Chapter 5. The fact that APA and CPA show quite different quenching behaviour, in spite of the same proximity of 2AP to a 5'G in both strands, supports the assertion that rapid ET from G is blocked by an intervening C[17], resulting in quenching by the nearest neighbour A or C becoming dominant.

4.4.1.2 450 nm Emission

It is notable that the decay parameters of the 2AP dimer in GPC differ from those in GPT. This suggests that 2AP may prefer to form a dimer with its 5' neighbouring base rather than with its 3' neighbour. If 2AP were to form a dimer with its 3' partner, the two neighbouring bases of the 2AP dimer in GPC and GPT would be the same, G and C on 5' and 3' sides, respectively, and might be expected to show similar fluorescence lifetimes, (assuming quenching interactions with nearest neighbour bases predominate). This is contrary to the observed behaviour.

4.4.2 Effect of Base-pairing

Table 4.5 shows a comparison of decay results for single strand GPC and a double strand GPC(C)[15] that has the same neighbouring bases of 2AP. The duplexes nomenclature is the same as that used for single strands with the base which 2AP is paired in the complementary strand given in brackets. The number and sequence of the bases are different between single and double strands, but the decay parameters of the duplex are typical of those in which 2AP is stacked directly with G. The magnitudes of τ_1 and τ_2 increase in the single strand compared to the duplex, especially, the shortest lifetime, which is twice as long in the single strand. This implies the π-stacking interaction of 2AP with G in the single strand is not as good which result in less efficient ET quenching. The magnitude of τ_4 decreases markedly in the single strand; the value of ~7 ns is significantly shorter than that of free 2APr in aqueous solution. This indicates that 2AP does not completely escape from inter-base interactions in the single strand, as it does in the duplex where this component is charateristic of extra-helical 2AP. In the single strand, there is no base-paring (hydrogen bonding) and therefore, there is an increase in the internal base motion and dynamic disorder compared with the double strand. Similar behaviour was observed by Ramreddy *et al*.[12]. The effect of removing base-pairing is also evident in the amplitudes of the decay components which differ significantly from those of the duplex. For 2AP in duplex DNA the values of the A factors almost invariably follow the trend shown by the present example, i.e. $A_1 >> A_2 > A_3 > A_4$. In the single strand, τ_1 remains the dominant component, but its amplitude is only about half that in the double strand, signifying a major reduction in the population of highly stacked states. There is a consequent increase in the amplitudes of the other three lifetimes. The increase in A_3 is particularly notable, indicating a large increase in the population of weakly stacked states.

Sequence	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
GPC single strand	0.08	0.59	2.4	6.9	0.45	0.16	0.31	0.08
GPC(C) double strand	0.04	0.45	2.6	10.3	0.83	0.09	0.05	0.02

Table 4.5 The global fluorescence decay parameters for single strand GPC at emission wavelength 385 nm and double strand GPC(C) at emission wavelength 390 nm, in solution at room temperature. Single strands are excited at 309 nm and double strands are excited at 317 nm. The data for double strand is reproduced from Neely *et al.*[15] result.

It can be concluded that base paring does affect the 2AP fluorescence behaviour through changing the π-stacking interactions between the intrastrand bases and the internal base motions. It seems that base-pairing increases the extent of the π-stacking interaction. It also decreases the internal base motion and restricts the conformational fluctuations of the DNA duplex structure.

The effect of base-pairing can also been seen in the behaviour of the 2AP dimer. The variation of the average lifetime of the 2AP dimer between the different single strand suggests that the quantum yield varies with different sequence contexts, but the data measured for 2AP dimer in DNA duplexes by Bonnist *et al.*[14] show little variation in quantum yield with sequence context. This suggests that 2AP dimer is less geometrically constrained in single strands leading to a greater variation in the degree of quenching. The average lifetime of the 2AP dimer in a duplex with sequence context APC(T) is about 1.45 ns[14] which is significantly shorter than that in the single strand CPA (3.91 ns). This is consistent with the conclusion that base-pairing increases the extent of π-stacking interaction, resulting in efficient quenching and decreasing the average fluorescence lifetime in the DNA duplex. If the emission intensities for single strands are scaled by the relative quantum yields, using the average lifetimes listed above, then the corrected emission intensities, as shown in Table 4.6, indicates the relative dimer populations for different single strands. It can be seen that there is significant variation among the five single strands. APA and GPC have almost the same dimer population; this is also the case for GPT and CPA. However, the explanation for this is still unclear.

Single strand	450 nm band relative intensity	Corrected 450 nm band relative intensity
APA	0.22	0.46
CPA	0.55	0.61
GPA	1	1
GPT	0.80	0.63
GPC	0.50	0.42

Table 4.6 The relative intensity of the 450 nm emission band and corrected 450 nm band relative intensity for each single strands after converted to the same relative quantum yield.

In order to compare the 2AP fluorescence decay behaviour at long wavelength between single and double strands, Table 4.7 compares the parameters of three single strands and one double strand[14] with similar sequence context. (The base length and full sequence are different between the single and double strands.) It can be seen that the magnitudes of τ_2, τ_3 and τ_4 increase in the single strands relative to the double strand. This is consistent with less π-stacking interaction in single strands than in double strands. τ_1 in some single strands is a little larger than the corresponding ones in double strands, which can be explained in the same way. However, τ_1 in some single strands is shorter than in double strands, which suggests that the higher level of internal base motion increases the non-radiative decay rate. Base-pairing increases the π-stacking interaction between the neighbouring bases, but decreases the internal base motion in double strand. Both π-stacking interaction and internal base motion influence the efficiency of quenching. If π-stacking interaction is the dominant factor, the lifetime will increase in going from duplex to single strand, whereas from Table 4.7, the amplitude of the longest lifetime increases in single strands compared with the double strand. This suggests that in the single strand it is easier for the dimer to escape interactions with its neighbouring bases, while remaining intact. This might be expected since the dimer would have to simultaneously break two base-pair bonds to adopt an extrahelical conformation.

Sequence	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
GPA single strand	0.15	1.1	4.3	15.1	0.39	0.24	0.14	0.23
GPT single strand	0.22	2.2	7.2	16.9	0.25	0.19	0.46	0.10
GPC single strand	0.14	1.4	5.7	15.0	0.31	0.22	0.24	0.23
GPG(T) double strand	0.19	0.97	3.6	9.2	0.53	0.26	0.16	0.05

Table 4.7 The global fluorescence decay parameters for single strands and double strands at emission wavelength 450 nm, in solution at room temperature. Single strands and double strands are excited at 360 nm. The data for double strand is reproduced from Bonnist *et al.*[14] result.

4.5 Conclusion

Overall, the fluorescence response of 2AP at 370 nm emission band in single strands is quite similar to that in double strands and described by four discrete components comparable conformational heterogeneity in both cases. The major conformation accounts for over 40% of the 2AP population in single strands and has a short lifetime of less than 120 ps. A small fraction, less than 10%, of the 2AP population has a lifetime of around 7 ns. Broadly speaking, the magnitudes of τ_1 and τ_2 increase and the magnitudes τ_3 and τ_4 decrease in single strands compared with those in DNA duplexes. The population of the shortest lifetime component decreases significantly and mainly transfers to the second and third components in single strands compared with double strands, resulting in the total population of the second and third components being larger than that of shortest lifetime component. Therefore, less quenched (less stacked) conformations are more highly populated in single strands. Hydrogen bonding increases the extent of the π-stacking interaction, decreases the internal base motion and restricts fluctuations in the DNA duplex structure. This has an important effect on the populations of the various conformational states. Overall, hydrogen bonding has an important influence on the 2AP fluorescence quenching process.

As in duplexes, 2AP in single strands shows a second emission band with maximum intensity at about 450 nm. This is because 2AP and one of its neighbouring bases form a ground-state heterodimer species. 2AP appears to preferentially form a dimer

with the natural base on its 5' side in the single strand. Dimer formation is less favourable in single strands than in DNA duplexes.

Like that in the DNA duplex, the fluorescence decay of 2AP dimer emission has four lifetimes that indicate the heterogeneity of the inter-base interactions of this species. The magnitudes of τ_2, τ_3 and τ_4 increase in single strands compared to those in double strands. Hydrogen bonding has similar effect on the fluorescence decay behaviour of the 2AP dimer and the 2AP monomer in single strands.

All the fluorescence lifetimes of the 2AP dimer are at least twice those of 2AP monomer, in single strands. Compared to the 2AP monomer, the 2AP dimer is geometrically constrained and π-stacking interaction between the 2AP dimer and its neighbours are weaker in single strands, resulting in slower quenching rates. The population differences between different conformational states are smaller for the dimer than the monomer implying that this too is influenced by π-stacking interaction.

4.6 References

(1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer: 1983.

(2) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I.; Hearst, J. E.; Wemmer, D. E.; Killman, P. A.; Turner, D. H. *Nucleic Acids: Structures, Properties, and Functions*; University Science Books; 1st edition: 2000.

(3) Protozanova, E.; Yakovchuk, P.; Frank-Kamenetskii, M. D. *J. Mol. Biol.* **2004**, *342*, 775-785.

(4) Hunter, C. A. *J. Mol. Biol.* **1993**, *230*, 1025-1054.

(5) Cozzi, F.; Cinquini, M.; Annunziata, R.; Dwyer, T.; Siegel, J. S. *J. Am. Chem. Soc.* **1992**, *114*, 5729-5733.

(6) Cozzi, F.; Cinquini, M.; Annuziata, R.; Siegel, J. S. *J. Am. Chem. Soc.* **1993**, *115*, 5330-5331.

(7) Schwabacher, A. W.; Zhang, S.; Davy, W. *J. Am. Chem. Soc.* **1993**, *115*, 6995-6996.

(8) Hunter, C. A. *Angew. Chem. Int. Edit.* **1993**, *32*, 1584-1586.

(9) Muehldorf, A. V.; Van Engen, D.; Warner, J. C.; Hamilton, A. D. *J. Am. Chem. Soc.* **1988**, *110*, 6561-6562.

(10) Cooper, V. R.; Thonhauser, T.; Puzder, A.; Schroder, E.; Lundqvist, B. I.; Langreth, D. C. *J. Am. Chem. Soc.* **2007**, *130*, 1304-1308.

(11) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142-10151.

(12) Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G. *J. Phys. Chem. B* **2007**, *111*, 5757-5766.

(13) Hardman, S. J. O.; Thompson, K. C. *Int. J. Quantum. Chem.* **2007**, *107*, 2092-2099.

(14) Bonnist, E. Y. M.; Jones, A. C. *ChemPhysChem* **2008**, *9*, 1121-1129.

(15) Neely, R. K., PhD thesis, The University of Edinburgh, **2005**.

(16) Bonnist, E. Y. M., PhD thesis, The University of Edinburgh, **2008**.

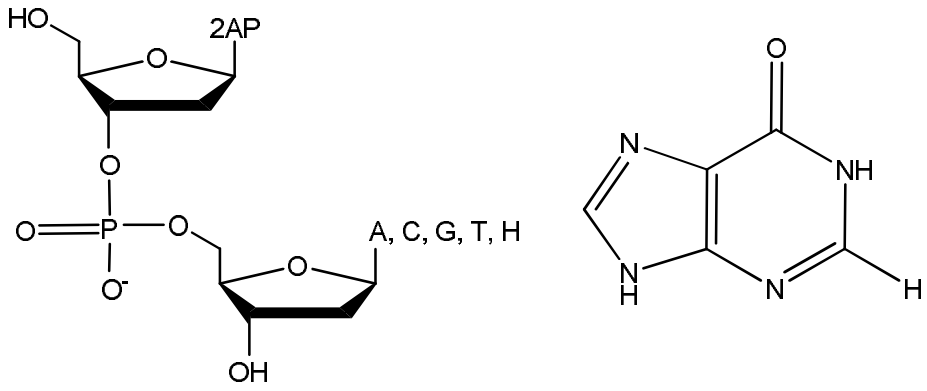
(17) Wan, C.; Fiebig, T.; Schiemann, O.; Barton, J. K.; Zewail, A. H. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 14052-14055.

Chapter 5

2-Aminopurine-Containing Dinucleotides

5.1 Introduction

The work reported in Chapter 4 shows the effect of hydrogen bonding on the conformations, π-stacking interactions and base mobility in DNA double and single strands. In order to compare the conformations, π-stacking interactions and base mobility in DNA single strands and dinucleotides, and investigate the importance of π-stacking interaction between a number of bases on maintaining the conformational behaviour and interactions in DNA, the fluorescence properties of five 2-aminopurine (2AP) - containing dinucleotides with the structure of 5' 2AP-X 3' with X = adenine (A), cytosine (C), guanine (G), thymine (T) and inosine (I), a simple model system, has been studied by steady-state and time-resolved fluorescence measurements, as reported in this chapter. In dinucleotides, the inter-base interactions of 2AP are limited to a single neighbouring base. The structures of the 2AP-containing dinucleotides are shown in Figure 5.1. Dinucleotides contain two sugars and one phosphate. 2AP is on the 5' side and the other natural base is on the 3' side of the dinucleotides used in this work. The structures of A, C, G, and T are illustrated in Figure 2.4 in Section 2.2.1 of Chapter 2.



(1) (2)

Figure 5.1 The structure of (1) the 2AP-containing dinucleotides and (2) hypoxanthine.

The dinucleotide 2AP-I has been included in the study because the calculated reduction potential of I is 1.5 V versus NHE[1], which is the same as the calculated reduction potential of 2AP*[1], indicating that I is redox inactive towards 2AP*. When it is incorporated in the 2AP-containing dinucleotide, it can serve as a reference for the electron transfer (ET) reactions between 2AP and the four natural bases. I is formed by attaching hypoxanthine (H), which is a derivative of purine, to a ribose ring. I is commonly present in tRNAs. The structure of H is shown in Figure 5.1.

Previously, the 2AP-containing dinucleotides have been studies theoretically and experimentally. Jean *et al.*[2] performed time-dependent density functional theory (TDDFT) with the B3LYP exchange-correlation functional and triple-ζ 6-311+G(d) basis set on the excited-state properties of 2AP-N dimers with N = A, C, G or T, which were constructed from canonical B - and A-form X-ray structures, with A being mutated to 2AP and the sugar and phosphate atoms being removed. (The structures of B - and A - forms of DNA are described in Section 2.2.1 of Chapter 2.) Each base was geometry-optimised separately at the MP2/6-31G(d,p) level of theory and then used to form the dimer starting structure. In order to keep the amino group in 2AP the same as it would be in a duplex structure, 2AP was minimized under C_s symmetry constraints. Their results showed that the neighbouring base had influence on the electronic structure and fluorescence of 2AP in dimers, as described below. Firstly, they predicted that whether 2AP stacked 5' or 3' to T made little difference to fluorescence characteristics. Therefore, their following calculations were only

performed on dimers with 2AP stacked on 5' side of the DNA natural base. Secondly, their results showed that in the constructed 2AP-T B-form dimer, the ground state (S_0) to the first singlet excited-state (S_1) absorption was extremely weak with the calculated oscillator strength 0.008. However, S_0 to the second singlet excited-state (S_2) absorption had a high oscillator strength, 0.094. Therefore, the low-lying S_1 of the 2AP-T B-form dimer is predicted to be a 'dark' state because of its very low oscillator strength, whereas S_2 is 'bright' state due to the high oscillator strength. They suggested that rapid nonradiative decay (internal conversion) would occur from S_2 to S_1, and S_1 would serve as a nonradiative sink for the loss of energy originally deposited in the S_2 state. As a consequence, they predicted the lifetime of 2AP in 2AP-T B-form dimer from the S_2 to S_0 transition would be dramatically shortened. Their prediction is consistent with experiment, in which there is a very short component in the fluorescence decay of 2AP-labelled DNA duplexes, as described in Section 2.2.2.3 of Chapter 2. Thirdly, in the constructed 2AP-T A-form dimer, the authors showed that the composition of the calculated electronic states was different to that of the B-form. They attributed this to the different relative orientations of the stacked bases between A - and B-form DNA, A-form having more extensive overlap of the six-membered ring. In the constructed 2AP-T A-form dimer, they showed that the calculated oscillator strength between the S_0 to S_1 and S_0 to S_2 absorptions were quite similar (~0.030). The predicted radiative lifetime of the vertical S_1 to S_0 transition was ~ 40 ns, which is significantly longer than the fluorescence lifetime of 2AP obtained from experiment ~10 ns, as described in Section 2.2.2.3 of Chapter 2. They suggested that non-radiative process would effectively compete to reduce the fluorescence quantum yield. Fourthly, they predicted the electronic character of 2AP-C (both A - and B-forms) dimer to be similar to that of 2AP-T dimer. Finally, they predicted that like the pyrimidine bases, the electronic character of the 2AP-A and 2AP-G B-form dimers are also similar. The predicted the radiative lifetimes of 2AP-A and 2AP-G B-form dimers to be 32 ns and 35 ns, respectively. They suggested fluorescence could arise from both S_2 to S_0 and S_1 to S_0 transitions with significantly low fluorescence quantum yield because of the mixed S_0 configuration.

Hardman *et al.*[3] performed CIS/3-21+G(d,p) computational calculations on double and single strand trinucleotide models (X2APX)·(YTY), where X and Y are the natural bases, in order to study the excited-state properties of 2AP in trinucleotides. The trinucleotides were created in both B - and A-forms, and the phosphate and sugar units were omitted from the calculation to keep the size of the molecule feasible for computational calculation. The structures of the canonical bases were used with those obtained from geometry optimisations at MP2/6-31G(d,p) level. The energies and oscillator strengths for vertical transitions to the excited-states of 2AP in (C2APC)·(GTG) were also performed using the TDDFT method at B3LYP functional and 3-21+G(d,p) basis set, with 2AP being in the optimised geometry of its first excited-state. The results from CIS/3-21+G(d,p) showed that the S_0 to S_1 fluorescence transition in all the cases with 2AP stacking with any of the natural nucleobases (no matter whether B - or A-form; double or single strand) involved an electron being promoted from an orbital focused mostly on 2AP to a vacant orbital also centred mainly on 2AP with $\pi\pi^*$ character. This indicated that there were no predicted dark states of charge transfer character lying below the S_0 to S_1 fluorescence transition, with the oscillator strength of the latter being no less than 0.100. This is contrary to the prediction by Jean *et al.*[2] that S_1 is a dark state. Although the first transition involved molecular orbitals that were mainly located on 2AP, there were still some components located on the other nucleobases of the same or the complementary strand to 2AP, showing that electrons were delocalized across 2AP and the neighbouring natural bases in trinucleotides. However, the results from TDDFT/3-21+G(d,p) for (C2APC)·(GTG) showed that the S_0 to S_1 fluorescence transition was charge-transfer in character with oscillator strength equal to 0. They attributed the discrepancy between CIS and TDDFT calculations to the incorrect form of the exchange-correlation B3LYP functional at long distances from the nucleus. As a consequence, they suggested that the multi-base systems could not be studied using the B3LYP functional.

Somsen *et al.*[4] studied the fluorescence lifetime 2AP-containing dinucleotides with 2AP at 5' side to each of the natural bases and I. The excitation wavelength was 290 nm. The fluorescence was collected through an emission filter (374.6 nm). The

instrument response function (IRF) (FWHM = 50 ps) was obtained from para-terphenyl in a 1:1 (volume) mixture of cyclohexane and CCl_4. Their results showed that 2AP in all of the dinucleotides had four fluorescence lifetimes in distinct regions: 18-35 ps, 0.3-0.8 ns, 1.8-3.1 ns and 8-9 ns, which were referred to as components 1 to 4. It was found that the 18-35 ps component, corresponding to the fast quenching, existed in all the dinucleotides including 2AP-I, in which I should be redox inactive. They suggested that the excited 2AP could be quenched by transition to a dark state, rather than by ET. The same authors also investigated the temperature dependence of the fluorescence decay[5]. With increasing temperature from 20 ºC to 60 ºC, the population of component 1 decreased and that of component 3 increased, but the sum of their populations did not change, indicating temperature dependent behaviour of components 1 and 3. The population of components 2 and 4 did not show much temperature dependence. They proposed that components 1 and 3 corresponded to a single ground-state species. They suggested a three ground-state and four excited-state model for these four fluorescence components of 2AP-containing dinucleotides, as described in Figure 5.2. Ground-states 2 and 4, corresponding to a single fluorescence decay component 2 and 4 respectively, each represents a ground-state species and has a single excited-state. Ground-state species X has two excited-states, X_1 and X_3, and shows a bi-exponential decay that corresponds to fluorescence lifetime components 1 and 3. They proposed that ground-state species X was mainly excited to excited-state X_1 with the relative excitation parameter $\alpha \approx 1$ (maximum $\alpha = 1$). The deexcitation of X_1 had two pathways: one was through fast quenching (with quenching rate q_1) and emitting fluorescence with lifetime corresponding to component 1; the other one was relaxation to the less quenched excited-state X_3 (with transition rate w_{13}). The quenching of this excited-state X_3 required reversal to excited-state X_1 (with transition rate w_{31}). They suggested that the highly quenched excited-state X_1 was possibly a stacked conformation of the dinucleotide and partially unstacked to state X_3 after excitation. The observed temperature dependence of the population of components 1 and 3 was explained by assuming that with increasing temperature, the deexcitation of X_1 was largely through relaxation to X_3 rather than fast quenching. They suggested their model was similar to the gating model proposed by O'Neil *et*

al.[6], as described in Section 2.2.2.2 of Chapter 2. They suggested that the difference between these two models was that the two excited-states X_1 and X_3 each corresponded to a ground-state species in the gating model[6]. They argued that the gating model could not explain the change of population X_1 and X_3 with increasing the temperature in their study.

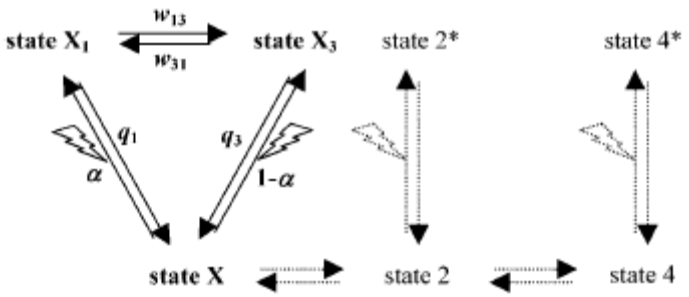


Figure 5.2 The three-state model of 2AP in dinucleotides. α and 1 – α represent the relative excitation of states X_1 and X_3, q_1 and q_3 indicate the quenching rates in these states, and w_{13} and w_{31} are the transition rates between the two. Reproduced from reference 5.

Later, Somsen *et al.*[7] suggested a much simpler model for the fluorescence decay of 2AP-containing dinucleotides, proposing that, the fluorescence decay of 2AP could be fitted by only two components: a slower stretched exponential component and a fast exponential decay component, using the equation below.

Equation 5.1 $$F(t) = a_s e^{-(t/\tau_s)^d} + a_f e^{-t/\tau_f}$$

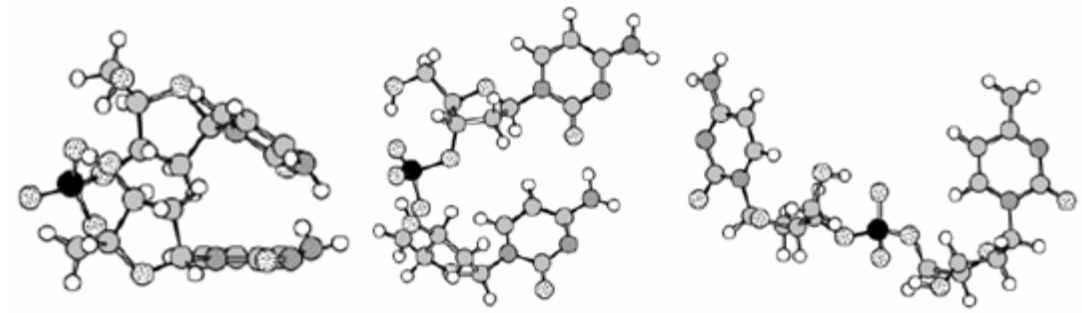
where, a_s and a_f are the amplitudes of the stretched and fast exponential respectively, τ_s and τ_f are the corresponding time constant, and d characterises the nature of the stretched component.

They suggested that the accuracy of the stretched-exponential fit was similar to that of a three-exponential fit through examining the residual plot of these two fits. They proposed that the former was a simpler representation of the data because of less degrees of freedom (5) compared to the latter (6). However, previously they had concluded that a three-exponential fit was inadequate, implying that the stretched exponential fit was inferior to the previously used four-exponetial decay. Their

results showed that the fast exponential component existed in all dinucleotides (A, C, G, T and I), of which the magnitudes (from 21 to 30 ps) were comparable to the fastest component obtained in four exponential fits shown in their previous papers[4,5]. This was assigned to quenching in a stacked conformation. They proposed that the ground-state equilibrium conformation of the bases was most favourable for the fluorescence quenching. The fast exponential component stood for quenching in the central part of the equilibrium distribution. The slow stretched exponential (from 0.63 to 2.2 ns) suggests a conformation that was out of equilibrium, of which the quenching was achieved after relaxation to the equilibrium conformation. The power value d of the stretched exponential is ~0.5 for purine bases and 0.7 for the pyrimidines and I, indicating the space for the conformational motion of the dinucleotide was restricted in the excited-state. According to Shlesinger *et al.*[8], d = 1/2 indicated relaxation in one dimension and d = 1 meant that in three dimensions. Therefore, Somsen *et al.* suggested that the conformational motion of 2AP dinucleotides with A and G had one degree of freedom, which was either the twisting or sliding, leading to the quenching conformation. However, the conformational motion of 2AP dinucleotides with T, C and I had a higher degree of freedom but less than three-dimensional space, such that both including twisting and sliding lead to the quenching conformation.

Gidden *et al.*[9] studied the gas-phase conformations and energies of 16 deprotonated dinucleotides by using ion mobility measurements and molecular modelling calculations. The arrival time distributions (ATD) for the dinucleotides showed multiple peaks representing the presence of multiple conformations[9]. The conformational family of each peak in the ATD was determined by the theoretical modelling of the dinucleotides through the AMBER 6.0 molecular mechanics/dynamics package. A 'scatter plot' of cross-section against energy of the structures was used to help determine the conformations observed in the ion mobility experiments. They found three different families of conformers in gas phase dinucleotides: stacked, H-bonded and open, as shown in Figure 5.3. Their results showed that in the purine-purine system (dAA, dAG, dGA and dGG), the bases in the stacked conformer are more parallel to each other, whereas in the purine-pyrimidine

and pyrimidine-pyrimidine systems, the bases are angled toward each, as shown in Figure 5.3. The dinucleotides in the stacked conformation were predicted to have lowest energy. The H-bonded conformer had similar energy to the stacked conformer, of which the energy was slightly smaller than that of open conformer. They proposed that stacking interaction between the bases was preferable in dinucleotides to hydrogen bonding due to the bases being anchored to the phosphate-sugar backbone and having a more limited range of movement. As a consequence, only one H-bond between the two bases was usually observed in the dinucleotides. Dinucleotides with 3' T prefer open conformation; those with 3' A prefer stacked conformation, those with 3' C favour H-bonded conformation. For those with 3' G, conformation depends on the 5' base; dAG and dCG prefer stacked conformation, dGG prefers the H-bonded conformation and dTG favours the open conformation. The reason why the 3' base has significant effect on the conformational preferences of the dinucleotides is not elucidated in this paper. At 80 K, dinucleotides showed multiple conformers. However at 300 K, the multiple conformers isomerised quickly and two conformers could interconvert quickly yielding a single conformer. At 500 K, the dinucleotides seemed to be in open conformation instead of stacked or H-bonded conformation. The results obtained from AMBER calculations showed that the 3' base did not have significant influence on the dinucleotide conformation, in disagreement with their experimental results. They attributed this to the AMBER calculations not considering electrons and suggested the use of higher-level of theory to investigate how the 3' base affects the conformational preference of these dinucleotides.



stacked H-bonded open

Figure 5.3 Three conformers in dinucleotides. Carbon atoms are gray, oxygen atoms are spotted, nitrogens are striped, hydrogens are white and the phosphorous atoms are black. Reproduced from reference 9.

Narayanan *et al.*[10] measured the steady-state fluorescence of the 2AP:native base monophosphate nucleotides (NMP) pairs including AMP, CMP, GMP and TMP. Through fitting the integrated steady-state fluorescence intensities for 2AP*:NMP pairs to the modified Stern-Volmer equation[11], shown in Equation 5.2, B, K_S and K_d values were obtained. K_S for 2AP:AMP pair (~ 5 M^{-1}) is at least twice that for the other pairs, suggesting AMP form complexes more readily with 2AP than the other NMPs.

Equation 5.2 $$\frac{I_0}{I} = B + (K_S + K_d)[Q] + K_S + K_d[Q]^2$$

where I_0 is the fluorescence intensity of 2AP, I is the fluorescence intensity of 2AP:NMP pair, B is the fitted y-intercept, K_S is the equilibrium constant for ground-state complexation, K_d is the constant for dynamic quenching and Q is the quencher.

Then, the fluorescence quenching constant k_q of 2AP* against various nucleotide quenchers were obtained by using the Equation 5.3. The fluorescence quenching constants from high to low are GMP (2.52×10^{-9} M^{-1} s^{-1}) > TMP (1.71×10^{-9} M^{-1} s^{-1}) > AMP (1.55×10^{-9} M^{-1} s^{-1}) > CMP (1.43×10^{-9} M^{-1} s^{-1}). Therefore, the fluorescence quenching experiments by modified Stern-Volmer analysis of steady-state emission

quenching on 2AP* showed that both purine and pyrimidine monophosphates could quench 2AP* with comparable efficiency. The order of quenching based on the fluorescence quenching constant is GMP > TMP > AMP > CMP.

Equation 5.3 $$K_q = \frac{K_d}{\tau_0}$$

where τ_0 is the fluorescence lifetime of 2AP, of which the average 11.4 ns is used by Narayannan *et al*.

They assumed that the fluorescence quenching of 2AP* by the nucleic acid monophosphates was through ET and then calculated the free energy of ET for both nucleobase oxidation and reduction by 2AP*. According to their free energy values, they suggested the thermodynamic driving force for ET from high to low is G > T $\approx$ A > C, which is consistent with the fluorescence quenching constants. In addition, they proposed that 2AP* oxidizes GMP and AMP but reduces TMP. The ET direction between 2AP*:CMP pair could occur either from 2AP* to CMP or from CMP to 2AP*.

In the present work, in order to investigate the nature of the conformational states that give rise to the heterogeneous decay of 2AP* and the role of base dynamics in populating these states, the fluorescence properties of the five 2AP-containing dinucleotides were investigated at 77 K, as well as a room temperature. Previously, Neely *et al*.[12] studied the fluorescence decay of three 2AP-labelled duplexes at 77 K. The three 2AP-labelled duplexes had sequence contexts: GPG(G), CPC(T) and TPA(T), where P stands for 2AP, the two bases beside 2AP are the two neighbouring bases and the base in the bracket is that paired with 2AP in the complementary strand. The fluorescence lifetimes of DNA duplexes in 10 M LiCl at room temperature and 77 K are shown in Table 5.1. Their results showed that in all the cases, the shortest decay component was eliminated at 77 K, and for TPA(T) the second shortest decay component also disappeared. Therefore, they proposed that rapid charge transfer quenching corresponding to the shortest decay component was due to base dynamics

and disappeared when the bases were frozen to 77 K and became static. It was concluded that the rapid quenched conformational structure was not a minimum energy geometry on the ground-state potential energy surface but a vibrationally excited-state in which the optimal stacked structure for rapid electron transfer was reached. GPG(G) and CPC(T) had three fluorescence lifetimes at 77 K instead of four fluorescence components at room temperature. However, TPA(T) only had two fluorescence components compared with four fluorescence lifetimes at room temperature. They attributed this to 2AP having no direct contact with G in the same or complementary strand and 2AP* in TPA(T) being inaccessible to ET from G and exhibiting a longer lifetime. 2AP in DNA duplexes having two or three fluorescence lifetimes suggested that the heterogeneity of the 2AP decay persisted at 77 K and DNA duplexes existed in a number of static conformational states, which corresponded to minima on the potential energy surface. When the duplexes were frozen to 77 K, the dynamic conformational population existing at room temperature was frozen to these static structures. The lifetimes τ_2, τ_3 and τ_4 had similar values at 77 K as at room temperature, indicating that the lifetimes of these conformational states were intrinsic. The large population of the shortest decay component at room temperature showed that a large proportion of duplexes (60% or more) reached this highly quenched conformation within picoseconds (or less) of excitation.

DNA duplex	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	Φ from decay
GPG RT	0.17	0.97	3.9	8.5	0.56	0.23	0.13	0.08	0.19
GPG 77 K	-	0.76	3.6	10.1	-	0.33	0.51	0.16	0.40
CPC RT	0.06	0.34	2.0	7.9	0.67	0.16	0.07	0.10	0.10
CPC 77 K	-	0.79	3.7	9.9	-	0.40	0.34	0.26	0.51
TPA RT	0.22	0.94	2.0	7.9	0.69	0.15	0.09	0.07	0.15
TPA 77 K	-	-	2.9	11.3	-	-	0.14	0.86	1.21

Table 5.1 Fluorescence decay parameters and relative quantum yields for duplexes GPG(G), CPC(T) and TPA(T) in 10 M LiCl at room temperature and 77 K excited at 320 nm. A factors are for 370 nm emission. Data is reproduced from reference 12.

In this chapter, the fluorescence spectra and lifetimes of 2AP-containing dinucleotides are presented at both room temperature and 77 K, and compared with

those of DNA single strands. The role of ET and base dynamics in quenching 2AP fluorescence in dinucleotides is elucidated.

5.2 Materials and Methods

Dinucleotides were dissolved in 0.01 M phosphate buffer (PBS) containing 0.0027 M KCl and 0.137 M NaCl, pH = 7.4 at 25 °C. To examine dinucleotides at 77 K, 10 M aqueous LiCl was used as the solvent because it forms a stable and reproducible transparent glass at 77 K[13] and has been shown not to significantly relatively perturb the DNA structure[14]. The absorbance (A) of the five 2AP-containing dinucleotides and 2AP riboside (2APr) in PBS buffer at 310 nm and LiCl at 305 nm are shown in table 5.2; these values were used to make the concentration correction to the fluorescence intensity.

Dinucleotide	A_{310} in PBS	A_{305} in LiCl
2AP-A	0.086	0.088
2AP-C	0.146	0.152
2AP-G	0.090	0.092
2AP-T	0.063	0.055
2AP-I	0.021	0.038
2APr	0.062	0.067

Table 5.2 The absorbance (A) of the five 2AP-containing dinucleotides and 2APr in PBS buffer at 310 nm and LiCl at 305 nm.

Steady-state fluorescence was measured using a Fluoramax spectrofluorometer, as described in Chapter 3. The excitation spectra of 2AP-containing dinucleotides in PBS buffer at room temperature were recorded with emission wavelengths from 360 nm to 500 nm, at 10 nm intervals, whereas the emission spectra were measured with excitation wavelengths from 290 nm to 400 nm, at 10 nm intervals.

The excitation and emission spectra of 2AP-containing dinucleotides in LiCl at room temperature were measured under the same condition to those in PBS buffer, with one more emission spectrum excited at 305 nm to calculate the fluorescence quantum

yields at the same excitation wavelength to those obtained from fluorescence lifetimes.

The method of freezing 2AP-containing dinucleotides to 77 K was described in Section 3.4 of Chapter 3. The excitation spectra of 2AP-containing dinucleotides in LiCl at 77 K were measured with emission wavelengths 305 nm, 310 nm, 320 nm, 350 nm, 360 nm and 370 nm, and the emission spectra were recorded with excitation wavelength 360 nm, 370 nm, 430 nm, 440 nm and 450 nm. All the spectra recorded in LiCl were corrected for background emission from the LiCl.

Quantum yields relative to free 2APr were calculated by integrating the intensity over the entire emission spectrum. Spectral intensity was calculated by integrating over a 20 nm wide band centred on the wavelength of the maximum emission intensity. Final values were converted to an equivalent concentration of 2APr.

Time-resolved fluorescence was recorded using the TCSPC technique as described in Chapter 3, at an excitation wavelength 305 nm. The fluorescence decays of dinucleotides in PBS buffer and LiCl at room temperature and 77 K were collected at three emission wavelengths: 360 nm, 380 nm and 400 nm. The fluorescence decays of 2APr were collected under to same condition as dinucleotides but at a single emission wavelength of 380 nm.

The relative quantum yield (Φ_{rel}) of the sample with respect to free 2APr in the same solvent was calculated from the fluorescence decay parameters according to equation 5.4. The quantum yield quoted is the average of the values at the three different emission wavelengths measured.

Equation 5.4 $$\Phi_{rel} = \frac{\langle \tau \rangle}{\tau_{2A\,\mathrm{Pr}}}$$

5.3 Results

5.3.1 Steady-state Fluorescence

5.3.1.1 In PBS Buffer at Room Temperature

The emission spectra of one of the dinucleotides, 2AP-A in PBS buffer, following excitation at 310 nm and 360 nm, are shown in Figure 5.4. Excitation at 310 nm generates an emission spectrum with maximum intensity at about 370 nm. Excitation at 360 nm produces a long-wavelength emission with maximum intensity at ~450 nm, but this emission is much weaker (~0.003) than the 370 nm emission spectrum.

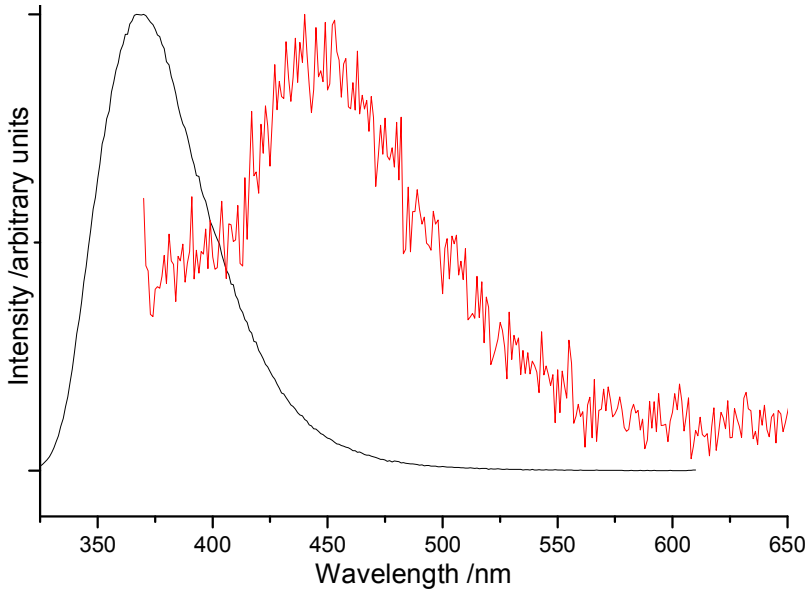


Figure 5.4 The emission spectra of 2AP-A in PBS buffer, excited at 310 nm (black line) and 360 nm (red line, intensity multiplied by a factor of ~350).

The fluorescence emission spectra of the other four 2AP-containing dinucleotides are very similar to those of 2AP-A and are not shown here.

The excitation spectra of all the five dinucleotides in PBS buffer have a peak at 305 nm with emission wavelength 370 nm. With emission wavelength 450 nm, the excitation spectra of all the five dinucleotides have some intensity at 360 nm, indicating the long wavelength emission is generated from a distinct ground state

species, although there is not a well-defined excitation peak because of the extremely low emission intensity of the dimer.

The 450 nm emission spectral properties are consistent with those in the DNA duplex[15] and single strands (in Chapter 4). The present observations confirm the previous conclusion that the heterodimer can be formed through π-stacking interaction between 2AP and any of the natural bases, and 2AP appears to preferentially form a dimer with its 5'side base rather than 3' side.

The intensity of each emission band relative to that of 2AP-I, and the intensity of the 450 nm emission band relative to that of the 370 nm emission band for dinucleotides in PBS buffer at room temperature is shown in Table 5.3.

Dinucleotide	370 nm band	450 nm band	Relative intensity (I_{450}/I_{370})
2AP-A	0.5	1.7	0.003
2AP-C	0.6	1.4	0.002
2AP-G	0.3	1.1	0.003
2AP-T	0.5	1.5	0.002
2AP-I	1	1	0.001

Table 5.3 The intensity of each emission band relative to that of 2AP-I, and the intensity of the 450 nm emission band relative to that of the 370 nm emission band for dinucleotides in PBS buffer at room temperature.

From Table 5.3, it can be seen that the 2AP-I has the maximum emission intensity at the 370 nm emission band, whereas it has the minimum emission intensity at the 450 nm emission band. Besides that, the variation in the short wavelength emission between the different dinucleotides is greater than that of the long wavelength emission band. The 450 nm emission band is much weaker (≤ 3‰) than the 370 nm emission band. Dinucleotides containing purines (A and G) have the similar relative intensity ratio between 450 nm and 370nm emission band, and this is also the case for the dinucleotides containing pyrimidines (C and T).

5.3.1.2 In LiCl at Room Temperature

The emission spectrum of one of the dinucleotides, 2AP-A in LiCl at room temperature, following excitation at 305 nm is shown in Figure 5.5. Excitation at 310 nm generates an emission spectrum with maximum intensity at about 367 nm. When it is excited at 360 nm or longer wavelength, there is no detectable fluorescence emission.

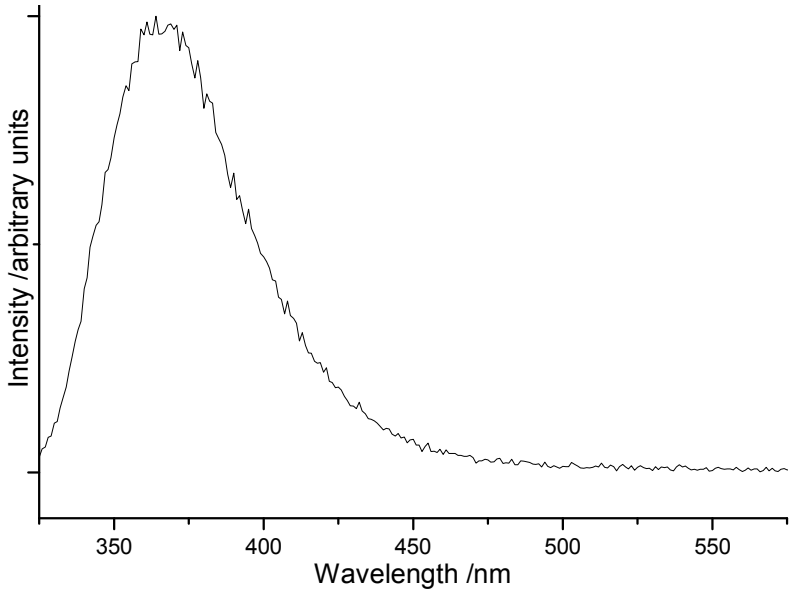


Figure 5.5 The emission spectrum of 2AP-A in LiCl at room temperature, excited at 305 nm.

The excitation spectrum of 2AP-A in LiCl at room temperature with emission wavelength 370 nm is illustrated in Figure 5.6. The excitation spectrum has a peak at 303 nm with emission wavelength 370 nm. There is no detectable fluorescence excitation with longer emission wavelengths.

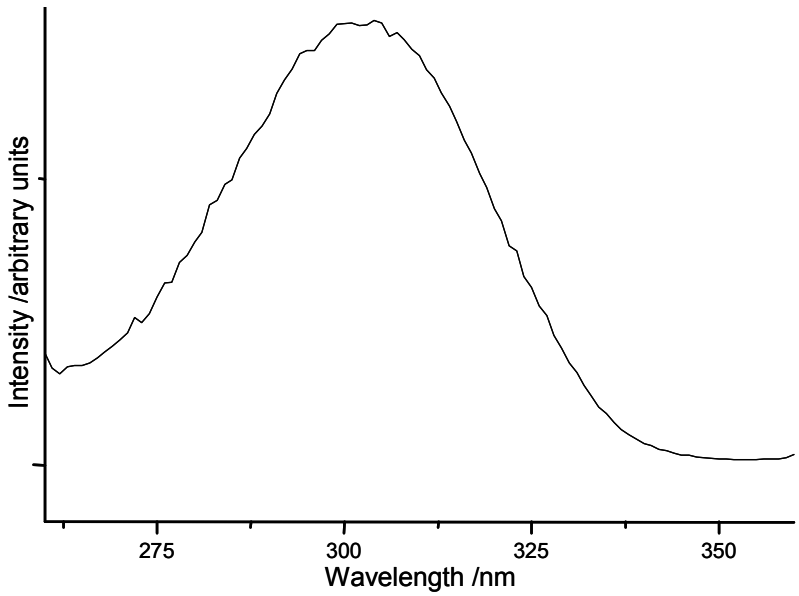


Figure 5.6 Excitation spectrum of 2AP-A in LiCl at room temperature with emission wavelength 370 nm.

The fluorescence excitation and emission spectra of the other four 2AP-containing dinucleotides are very similar to those of 2AP-A and are not shown here. There is no detectable long wavelength emission of 2AP-containing dinucleotides in LiCl, suggesting that the high concentration of LiCl inhibits dimer formation or enhances quenching of dimer emission.

5.3.1.3 In LiCl at 77 K

The emission spectrum of one of the dinucleotides, 2AP-A in LiCl at 77 K, following excitation at 305 nm is illustrated in Figure 5.7. Excitation at 305 nm generates an emission spectrum with maximum intensity at about 363 nm. Excitation at 360 nm appears to produce a very weak emission with maximum intensity at about 400 nm, but the intensity was too weak compared with background LiCl emission to record a reliable spectrum. The observation of weak dimer emission at 77 K suggests that the absence of dimer emission in LiCl at room temperature is because of enhanced quenching rather than inhibition of dimer formation.

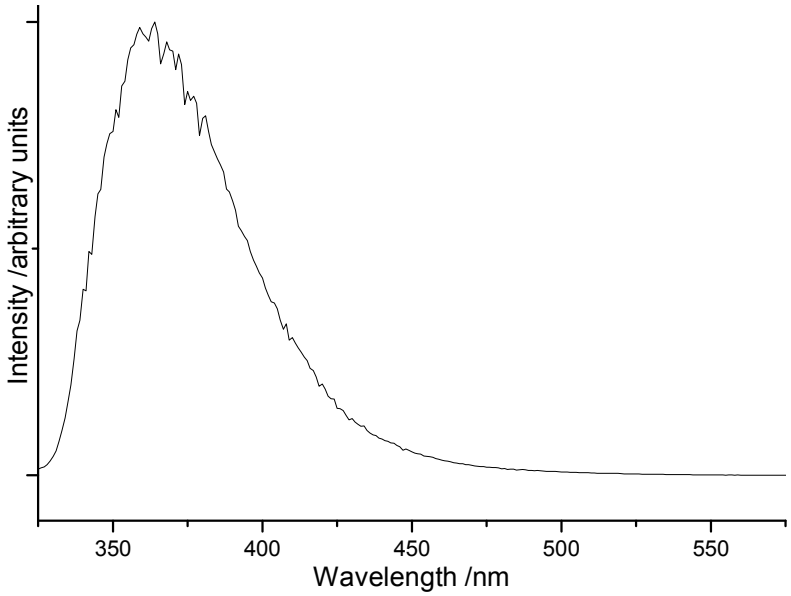


Figure 5.7 Emission spectra of 2AP-A in LiCl at 77 K, excited at 305 nm.

The excitation spectrum of 2AP-A in LiCl at 77 K with emission wavelength 370 nm is shown in Figure 5.8. The excitation spectrum has a peak at 317 nm with emission wavelength 370 nm. With emission wavelength 430 nm, the excitation spectrum has some intensity at 360 nm, indicating the long wavelength emission is generated from a distinct ground state species, although there is not a well-defined excitation peak because of the extremely low emission intensity of the dimer.

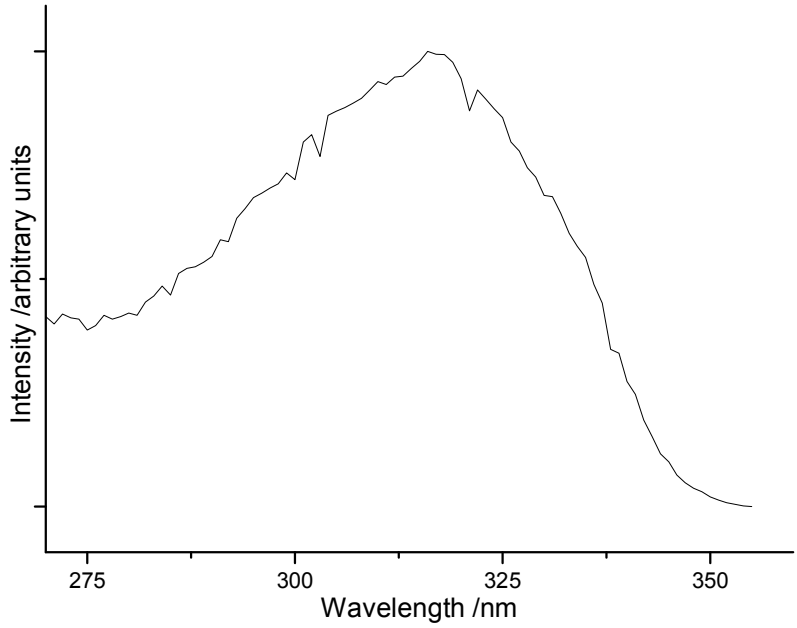


Figure 5.8 Excitation spectrum of 2AP-A in LiCl at 77 K with emission wavelength 370 nm.

Except for 2AP-I, which does not have long wavelength emission with excitation wavelength 360 nm, the fluorescence excitation and emission spectra of the other three 2AP-containing dinucleotides are very similar to those of 2AP-A and are not shown here. The 400 nm emission spectral properties are similar to those of DNA duplexes measured previously by Bonnist *et al.*[15] with the maximum intensity blue-shifted ~20 nm, suggesting the excitation energy of the dimer in dinucleotides is slightly greater than that in DNA duplexes.

The intensity of the 363 nm emission band relative to that of 2AP-I for dinucleotides is represented in Table 5.4. 2AP-G has the lowest emission intensity and 2AP-I has the highest emission intensity among all the dinucleotides. The other three dinucleotides have similar, intermediate emission intensity.

Dinucleotide	363 nm band
2AP-A	0.8
2AP-C	0.8
2AP-G	0.6
2AP-T	0.8
2AP-I	1

Table 5.4 The intensity of emission peak 363 nm for four dinucleotides in LiCl relative to that of 2AP-I.

The intensity of the 363 nm emission band at 77 K relative to that at room temperature for all the five dinucleotides and 2APr in LiCl are shown in Table 5.5. The fluorescence intensity of 2APr at 77 K is 6 times as much as that at room temperature, as a result of elimination of solvent relaxation. The ratios of fluorescence intensity between 77 K and room temperature are very similar among the three dinucleotides 2AP-G, T, and I. The ratios for 2AP-A and 2AP-C, which are also similar, are slightly smaller than that of the other three dinucleotides. If 2AP-containing dinucleotides have similar interaction with solvent compared with 2APr, the fluorescence intensity of 2AP-containing dinucleotides will also increase 5 times at 77 K. However, the total intensity of 2AP-containing dinucleotides increases roughly 40 times, indicating the other 35 times enhancement of the fluorescence intensity of dinucleotides at 77 K is due to removing quenching through base

dynamics. It also indicates that the quenching of fluorescence intensity of 2AP in dinucleotides is mainly from the base dynamics, and the static quenching only plays a small effect on it, as described in DNA duplexes[12].

Dinucleotide	$I_{77\ K}/I_{RT}$
2AP-A	38
2AP-C	39
2AP-G	42
2AP-T	44
2AP-I	43
2APr	6

Table 5.5 The intensity of the 363 nm emission band at 77 K relative to that at room temperature for all the five dincleotides and 2APr in LiCl.

5.3.2 Fluorescence Lifetimes

5.3.2.1 In PBS Buffer at Room Temperature

Fluorescence decay parameters at excitation wavelength 305 nm are represented in Table 5.6. Amplitude parameters (A) are for emission wavelength 380 nm. The full set of the decay parameters for the five dinucleotides in PBS buffer at all three emission wavelengths are given in Tables A13 to A17 in Appendix A.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	Φ from spectrum	Φ from decay
2AP-A	-	0.61	2.0	9.6	-	0.64	0.33	0.03	0.10	0.14
2AP-C	0.16	0.78	2.4	9.2	0.17	0.14	0.68	0.01	0.12	0.19
2AP-G	0.06	0.63	1.8	10.3	0.47	0.14	0.33	0.06	0.06	0.13
2AP-T	0.07	0.65	2.3	10.2	0.27	0.11	0.59	0.03	0.10	0.17
2AP-I	-	0.83	3.1	8.5	-	0.33	0.61	0.06	0.19	0.27
2APr	-	-	-	9.8	-	-	-	1	1	1

Table 5.6 The global fluorescence decay parameters for dinucleotides and 2APr excited at 305 nm. A factors are for 380 nm emission. The quantum yields of dinucleotides relative to 2Apr, calculated from both emission spectra and fluorescence decay parameters, are also shown. Experimental uncertainties (estimated by dividing standard deviation of values by the corresponding values): τ_1 (≤ 14%), τ_2 (≤ 6%), τ_3 (≤ 1%), τ_4 (≤ 2%), A_1 (≤ 6%), A_2 (≤ 3%), A_3 (≤ 1%), A_4 (≤ 6%) and quantum yields (≤ 7%).

It can be seen that fluorescence decay of 2AP in dinucleotides 2AP-A and 2AP-I is characterised by three discrete components, and that in the other three dinucleotides it is characterised by four discrete components, indicating that the existence of different conformations in dinucleotides as observed in previous studies of 2AP-labelled DNA duplexes (as described in Section 2.2.2.3 of Chapter 2) and single strands (as described in Chapter 4). The most significant difference among theses dinucleotides is the variation of the shortest lifetime component τ_1 and the identity of the dominant component. The shortest lifetime component is missing in 2AP-A and 2AP-I, similar to the observation for single strand APA in Chapter 4. The shortest lifetime component in 2AP-G and 2AP-T is almost identical ~60 ps, whereas that in 2AP-C is much longer (~160 ps). Broadly speaking, the other three lifetimes τ_2, τ_3 and τ_4 have similar values, 0.7 ns, 2.3 ns and 9 ns, for all the five dinucleotides. The dominant component in 2AP-G is the first lifetime component which accounts for ~50% of the dinucleotide population. The dominant component in 2AP-A is the second lifetime component (64%), the population of which is similar to the sum of the first two component populations in 2AP-G. The population of τ_3 and τ_4 in 2AP-A are almost identical to the corresponding ones in 2AP-G. The dominant component in 2AP-C, 2AP-T and 2AP-I is the third lifetime component that is no less than 59%. The population of each lifetime component in 2AP-C is very similar to the corresponding ones in 2AP-T. Less than 6% of the 2AP molecules in all the dinucleotides have fluorescence lifetimes about 9 ns that is similar to the fluorescence lifetime of 2APr, indicating that only a very small fraction of 2AP is free from π-stacking interactions in the dinucleotides. There is almost no dependence of the amplitudes of the lifetimes on different emission wavelength (see Tables A13 to A17 in Appendix A), indicating that the decay components have very similar emission spectra.

5.3.2.2 In LiCl at Room Temperature

When the fluorescence lifetimes of 2AP-containing dinucleotides were measured in LiCl at room temperature, it was found that the fluorescence count rate (intensity) increased slowly over a period of three hours. To investigate the effect, the

steady-state spectra and fluorescence decays were measured repeatedly at different start time: 0 minuete (min), 35 mins, 70 mins, 104 mins, 140 mins and 180 mins. 2AP-containing dinucleoitdes in LiCl were made by diluting the stock solution stored in PBS buffer ten times using LiCl. They were then transferred to cuvette and put in the sample chamber. The whole process was finished within 30 seconds; therefore, start collecting time 0 min means that the measurements began immediately after the dinucleotides in LiCl were made. The decays collected below 180 mins were at a single emission wavelength 380 nm, whereas those collected at 180 min were at three emission wavelengths 360 nm, 380 nm and 400 nm

Fluorescence decay parameters for dinucleotides excited at 305 nm and collected at 180 mins are shown in Table 5.7. Amplitudes (A) in Table 5.7 are for emission wavelength 380 nm. The full set of the decay parameters are presented in Tables A18, and A20 to A23 in Appendix A.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	Φ from spectrum	Φ from decay
2AP-A	0.06	0.84	3.8	7.5	0.63	0.13	0.17	0.07	0.10	0.18
2AP-C	0.06	0.69	3.9	7.0	0.69	0.05	0.22	0.04	0.10	0.17
2AP-G	0.07	0.59	3.7	7.7	0.71	0.09	0.13	0.07	0.07	0.15
2AP-T	0.05	0.60	3.7	5.7	0.70	0.06	0.14	0.10	0.09	0.15
2AP-I	0.06	0.96	3.2	6.0	0.43	0.17	0.20	0.20	0.12	0.27
2APr	-	-	-	7.5	-	-	-	1	1	1

Table 5.7 The global fluorescence decay parameters for dinucleotides which are excited at 305 nm and collected at 180 mins. A factors are for 380 nm emission. The relative quantum yields calculated from both spectra and fluorescence fit parameters are also shown. Experimental uncertainties (estimated by dividing standard deviation of values by the corresponding values): τ_1 and τ_2 (≤ 5%), τ_3 (≤ 3%), τ_4 (≤ 2%), A_1 (≤ 2%), A_2 (≤ 3%), A_3 (≤ 6%), A_4 (≤ 10%) and quantum yields (≤ 17%).

The fluorescence decay of 2AP in all the dinucleotides required four discrete components to give a satisfactory fit. As shown in Table A19 and A24 in Appendix A, the decays of 2AP-A and 2AP-I, cannot be fitted adequately with three components, unlike their decays in PBS. The shortest lifetime component exists in all the dinucleotides in LiCl including 2AP-A and 2AP-I, and is the dominant component

with population over 60% except for 2AP-I (~40%), indicating the shortest component and dominant component do not show the sequence-dependence seen in PBS buffer. In general, there is very little variation of each fluorescence lifetime, which are ~60 ps, ~0.7 ns, ~3.7 ns and ~7.0 ns, among the dinucleotides. In addition, the variation in population of the decay components is smaller than that in PBS buffer, indicating LiCl solvent decreases the difference of fluorescence decay behaviour of 2AP in different dinucleotides. The magnitude of the longest lifetime is a little shorter than that in PBS buffer, consistent with 2AP experiencing a different solvent environment.

The amplitudes of the lifetimes show a dependence on emission wavelength, indicating that the decay components have different emission spectra. The amplitudes of the lifetimes at three emission wavelengths of 2AP-A are shown in Table 5.8. Those of the other four dinucleotides are similar to 2AP-A, and are shown in Tables A20 to A23 in Appendix A. It can be seen that with emission wavelength, from 360 nm to 400 nm, the amplitude of τ_1 increases noticeably in all the dinucleotides, showing that the emission spectrum of this highly stacked conformation is red-shifted compared to other conformations, whereas the amplitudes of the other three lifetime components decrease, indicating the emission spectra of the imperfectly stacked and unstacked conformations are blue-shifted relative to the highly stacked conformation. This behaviour differs from the wavelength-independent amplitudes seen in PBS buffer. Neely *et al.*[12] also showed the wavelength dependence of amplitudes of lifetimes in certain DNA duplexes. However, the amplitude of τ_1 was found to decrease and the amplitudes of the other three lifetimes to increase with increasing the emission wavelength[12], the opposite trend to that seen here in dinucleotides.

Emission Wavelength/ nm	A_1	A_2	A_3	A_4
360	0.55	0.16	0.2	0.09
380	0.63	0.13	0.17	0.07
400	0.68	0.12	0.14	0.06

Table 5.8 A factors of 2AP-A in LiCl at three emission wavelengths.

Fluorescence decay parameters for dinucleotides excited at 305 nm and collected at 0 min are shown in Table 5.9. Fluorescence decay parameters for dinucleotides excited at 305 nm and collected at 35 mins, 70 mins, 104 mins, 140 mins are shown in Tables A25 to A28 in Appendix A.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	Φ from spectrum	Φ from decay
2AP-A	0.05	0.43	2.8	6.3	0.84	0.05	0.04	0.07	0.029	0.08
2AP-C	0.05	0.33	3.2	5.6	0.84	0.03	0.04	0.08	0.017	0.09
2AP-G	0.05	0.26	2.6	6.4	0.80	0.14	0.02	0.04	0.015	0.05
2AP-T	0.05	0.49	2.3	5.7	0.82	0.04	0.02	0.12	0.016	0.11
2AP-I	0.05	1.0	3.2	6.0	0.62	0.11	0.12	0.15	0.035	0.19

Table 5.9 The global fluorescence decay parameters for dinucleotides which are excited at 305 nm and collected at 0 min. A factors are for 380 nm emission. The relative quantum yields calculated from both spectra and fluorescence fit parameters are also shown. Experimental uncertainties (estimated by dividing standard deviation of values by the corresponding values): τ_1 (≤ 3%), τ_2 (≤ 10%), τ_3 (≤ 7%), τ_4 (≤ 1%), A_1 (≤ 1%), A_2 (≤ 13%), A_3 (≤ 7%), A_4 (≤ 3%) and quantum yields (≤ 17%).

Comparing the fluorescence lifetimes of 2AP in dinucleotides collected from 0 min to 180 mins, all the four fluorescence lifetimes in the dinucleotides increase with increasing the time. Especially with increasing time, the population of the shortest lifetime component (A_1) decreases more than 10% and mainly transfers to the third component (A_3), the trend of which is similar to that with increasing temperature measured by Somsen *et al.*[5], as described in the introduction in Section 5.1. Therefore, the fluorescence intensity of 2AP-containing dinucleotides rises with increasing the time and is stable after 3 hours. The equilibrium time of dinucleotides in LiCl is much longer than in PBS buffer probably due to the significantly higher viscosity of the 10 M LiCl.

5.3.2.3 In LiCl at 77 K

Fluorescence decay parameters at excitation wavelength 305 nm and emission wavelength 360 nm are presented in Table 5.10. The decay parameters were obtained by fitting an individual decay instead of global analysis of three decays. This is

because three components were needed to give a good global fit, whereas only two components were required to fit individual decays, with only small variation in the fitted lifetimes at the three emission wavelengths. The failure of global analysis may be because the quality of the decay data at 77 K is not as high as that at room temperature and there are discrepancies between decays collected at the three different emission wavelengths, or there may be genuine slight variation of lifetime with emission wavelength. The results of global fitting of the decays of 2AP-A with three or two components are shown in Tables A29 and A30 in Appendix A. The parameters obtained by individual fitting of the same three decays are in Tables A31 and A32 in Appendix A.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	Φ from spectrum	Φ from decay
2AP-A	-	-	4.5	14.1	-	-	0.11	0.89	0.7	1.6
2AP-C	-	-	3.2	14.1	-	-	0.05	0.95	0.7	1.6
2AP-G	-	-	3.8	14.0	-	-	0.16	0.84	0.5	1.5
2AP-T	-	-	3.2	14.5	-	-	0.06	0.94	0.7	1.7
2AP-I	-	-	4.0	15.0	-	-	0.11	0.89	0.9	1.7
2APr	-	-	-	8.4	-	-	-	1	1	1

Table 5.10 The fluorescence decay parameters for dinucleotides and 2APr excited at 305 nm and at emission wavelength 360 nm. The relative quantum yields calculated from both spectra and fluorescence fit parameters are also shown. Experimental uncertainties (estimated by dividing standard deviation of values by the corresponding values): τ_3 (≤ 4%), τ_4 (≤ 1%), A_3 (≤ 2%), A_4 (< 1%) and quantum yields (≤ 5%).

All the dinucleotides have two discrete fluorescence lifetime components, showing that freezing the dinucleotides at 77 K has a significant effect on the decay function, as observed in previous studies of 2AP-labelled duplexes[12]. In all cases, the first two shortest decay components have vanished. The two lifetimes have similar values, ~ 4.0 ns and ~14 ns, for all the five dinucleotides. The lifetime τ_3 at 77 K is similar to the corresponding one at room temperature, suggesting this lifetime measured at room temperature is the intrinsic lifetime of a particular conformational state, as seen in 2AP-labelled duplexes[12]. The magnitude of τ_4 is nearly twice that of 2APr. The fluorescence decays are dominated (> 83%) by the ~14 ns component.

5.4 Discussion

5.4.1 Discrepancy between Relative Quantum Yield Determined from Spectra and Decay Parameters

Table 5.11 shows the ratio of the relative quantum yield determined from the decay parameters to that determined from the emission spectrum for each of the dinucleotides, under the different solution conditions. It can be seen that, in all cases, the relative quantum yield calculated from the decay parameters is considerably greater than that calculated from the spectra for dinucleotides, typically by a factor of about 2. A similar discrepancy was reported previously for DNA duplexes by Bonnist *et al*.[16]. This discrepancy can be accounted for by the existence of 2AP species with fluorescence lifetimes too short to be detected by the TCSPC technique used here, that is to say, lifetimes less than ~10 ps. Within the limitations of the present experiments, these species are effectively non-emissive. The existence of such species has been demonstrated by Wan *et al*.[17] who investigated fluorescence lifetimes of 2AP-A and 2AP-G dimers in solution with sub-picosecond time resolution, using both femtosecond transient absorption and fluorescence upconversion tenchniques. They found that ~50% of initial fluorescence intensity of the 2AP-containing dimers was lost in less than 200 fs (the limit of their time resolution). They attributed this extremely rapid non-radiative decay to barrierless charge transfer of vibrationally hot molecules prior to vibrational relaxation. Their observation that ~50% of the excited 2AP population is non-emissive is consistent with the result here that the quantum yield obtained from the decay parameters is roughly twice that measured from spectra.

Dinucleotide	PBS, RT	LiCl, RT	LiCl, 77 K
2AP-A	1.4	1.8	2.3
2AP-C	1.6	1.6	2.2
2AP-G	2.2	2.0	2.8
2AP-T	1.7	1.7	2.4
2AP-I	1.4	2.2	1.9

Table 5.11 The quantum yield from fluorescence fit parameters relative to that from steady-state spectra for five dinucleotides in PBS buffer and LiCl solution at room temperature and 77 K.

5.4.2 Comparison with Previous Literature Measurements

The decay parameters obtained in the present work for the dinucleotides in PBS buffer at room temperature differ significantly from those reported previously by Somsen *et al*.[4]. In the latter work, all the five dinucleotides, including 2AP-I, were reported to have a very short lifetime component ~30 ps, whereas in the present work the very short component is not seen for 2AP-A and 2AP-I. In addition, the shortest lifetime component in the other three dinucleotides measured in the present work is considerably longer than the values reported by Somsen.

There are some significant differences between the experimental conditions used in the two studies. In the present work, the excitation wavelength used was 305 nm, whereas Somsen *et al*.[4] used 290 nm. Beside that, Somsen *et al*. did not use a monochromator to spectrally resolve the emission, but collected fluorescence through a bandpass filter. It is possible, therefore that scattered excitation light could reach the detector and be manifested as a very short decay component. In the present work both a bandpass filter and a monochromator were used to insure exclusion of scattered light. In order to check whether the discrepancy between this work and that of Somsen could be due to the use of different excitation wavelengths, the fluorescence decay of 2AP in four dinucleotides (excluding 2AP-I) in PBS buffer at room temperature were measured at excitation wavelengths 292 nm, 276 nm and 261 nm. The full set of decay parameters are shown in Tables A33 to A35 in Appendix A. With decreasing excitation wavelength from 305 nm to 261 nm, 2AP-A does show an additional component with lifetime as short as ~0.3 ns, and the shortest lifetime

component of 2AP-C decreases to 80 ps, but they are still considerably longer than those obtained by Somsen *et al.*[4]. The shortest lifetime component in 2AP-G and 2AP-T is ~40 ps, which is slightly longer than the corresponding one obtained by Somsen *et al.*[4]. It is notable that in Somsen's work the quantum yields calculated from the decay parameters in all the five dinucleotides are very similar to the corresponding ones calculated from the spectra, in fact the former are slightly smaller. This is not consistent with the expected quantum yield discrepancy between the spectral and decay data, as discussed above, and casts doubt on the validity of the short decay components reported by Somsen *et al.*[4]. There is generally good agreement between the spectrally derived quantum yields reported by Somsen and those measured in the present experiments.

Wan *et al.*[17] also obtained a very short lifetime component of 80 ps for 2AP-A using the fluorescence upconversion technique. However, the concentration of 2AP-A in Wan' work is 1 mM, which is ~ 100 times greater than in this work. At this high concentration, aggregation of the dinucleotides is likely, leading to similar effects as those which result in observation of a short decay component in concentrated LiCl as discussed in Section 5.4.5, below.

5.4.3 Electron Transfer Quenching in PBS at Room Temperature

There is strong evidence that the dominant mechanism for quenching of 2AP in duplex DNA is ET from G, which seems also the dominant mechanism for quenching of 2AP in 2AP-G dinucleotide. It will now be considered whether this holds true in the other four dinucleotides.

The reduction potentials of the natural bases, versus NHE,[18] are: of A 1.42 V, of G 1.29 V, of C ~1.6 V, of T ~1.7 V. The reduction potentials of both $2AP^*$ and I are 1.5 V versus NHE[1]. The driving force for ET between 2AP* and other bases, corresponding to the difference of reduction potential between 2AP* and the other

bases, from high to low is: G (0.21 V versus NHE) > T (0.2 V versus NHE) > C (0.1 V versus NHE) > A (0.08 V versus NHE) > I (0 V versus NHE).

The relative quantum yield calculated from the emission spectra of 2AP-containing dinucleotides from low to high is G < A ≈ T < C < I. Therefore the quenching efficiency of 2AP-containing dinucleotides from high to low is G > A ≈ T > C > I, which correlates with the driving force for ET, except that 2AP-A is out of sequence. Although ET between I and 2AP* is not expected, it is evident that interaction with I quenches 2AP fluorescence. The imperfect correlation between quantum yield and ET driving force could be due to firstly, the change of the redox potentials of the individual bases upon incorporation into dinucleotides and secondly, the influence of the unknown activation energies for the ET reactions[19]. However, the trend of quenching efficiency of 2AP-containing dinucleotides correlates with the thermodynamic driving force for ET in 2AP:native base monophosphate nucleotides (NMP) pairs, including AMP, CMP, GMP and TMP, in aprotic organic solvents (acetonitrile or *N,N*-dinetylformamide)[10]. Although, the thermodynamic driving force for ET in 2AP:native base pairs would change in aqueous solution. It appears, therefore, that the mechanism of quenching of 2AP* in the other four dinucleotides involves ET, in a similar way to that in 2AP-G, which is consistent with the conclusion of Larsen *et al.*[19] that the fluorescence quenching of 2AP has at least partly a charge-transfer process at its base. From the reduction potential, it is proposed that for the purines ET is from G or A to 2AP*, whereas for the pyrimidines ET is from 2AP* to T and C. ET between 2AP* and I is probably dual-direction either from 2AP* to I or from I to 2AP* due to the same reduction potential between 2AP* and I.

The magnitude of the shortest lifetime component for dinucleotides from low to high, is 2AP-G < 2AP-T < 2AP-C, correlates with the driving force for ET. This confirms the assignment of the rapid non-radiative decay channel to ET quenching. The shortest lifetime component is absent from 2AP-I which is consistent with the absence of rapid ET in this case. The amplitude (A_1) of the shortest component also correlates with the facility of ET. A larger fraction of the conformational population

is amenable to rapid ET in 2AP-G, whereas in 2AP-T and 2AP-C the requirement to adopt a highly stacked conformation to facilitate fast ET is more critical and is satisfied by a smaller fraction of the population.

2AP-A appears to be anomalous in lacking the shortest lifetime component, but this was also the case for the APA single strand reported in Chapter 4 and for the 30 adenine (poly A) containing single strand measured by Ramreddy *et al.*[20]. This suggests the fast ET does not occur between A and 2AP* in any of these systems. The ET is governed by three factors: the redox potential difference between the electron donor and acceptor, π-stacking interactions and base dynamics (motions) (as described in Section 2.2.2.2 of Chapter 2). Therefore, it is proposed that 2AP and A cannot form the necessary π-stacking interaction to facilitate the fast ET quenching. The different π-stacking interaction behaviour of 2AP-A to the other dinucleotides might be due to the structural similarity between A and 2AP. When 2AP is stacked with A in the dinucleotide, the exocyclic amine group in both 2AP and A might hinder the formation of a highly π-stacked conformation.

Comparing the magnitudes of τ_2 and τ_3 in the four dinucleotides, excluding the anomalous 2AP-A, shows that these lifetimes also correlate inversely with the driving force for ET, indicating that the slower quenching in the less well stacked conformations also involves ET.

5.4.4 Conformational Properties Compared with Single Strands

In both dinucleotides and single strands, the magnitude of τ_1 depends strongly on the sequence context. The efficient quenching of 2AP* by a neighbouring G is results in similar values of τ_1, around 60 ps, in both systems. The magnitude of τ_2 is slightly greater in dinucleotides than in single strands, while that of τ_3 is similar. This implies that the π-stacking interaction in dinucleotides is similar to that in single strands and

is affected little by the multiplicity of π-stacking interactions present in single strands. It also suggests that the imperfectly π-stacked conformations in dinucleotides are similar to those in single strands. The magnitude of the longest lifetime component of 2AP in dinucleotides is longer than that in single strands, suggesting that 2AP is able to completely escape from inter-base interactions in dinucleotides. It is proposed that internal base motion and dynamic disorder in dinucleotides is greater than in single strands because of the freedom from the constraints of multiple inter-base stacking interactions. Therefore, the 2AP in dinucleotides has a greater probability of reaching the totally unstacked conformation.

The effect of removing the series of π-stacking interactions is also evident in the amplitudes of the decay components of the dinucleotides, which differ significantly from those of the single strands. The dominant component for all the dinucleotides, except for 2AP-G, is τ_2 or τ_3, whereas the dominant component is τ_1 in the single strands, implying a major decrease in the population of highly stacked states in the former. In general, $A_1$and A_4 decrease and transfer to A_2 and A_3 in dinucleotides, causing the sum of A_2 and A_3 to be more than 70% in dinucleotides, which is much greater than in single strands. This indicates a significant rise in the population of weakly stacked states. Thus, it seems that the series of base π-stacking interactions decreases the internal base motion and limits the conformational fluctuations of the single strand structure.

5.4.5 The Effect of LiCl

The introduction of the dinucleotides into 10 M LiCl (at room temperature) has a profound effect on their fluorescence decay, giving rise to a very short, ~60 ps component in all cases, including 2AP-A and 2AP-I which show no short component in PBS. The LiCl solvent appears to enhance the inter-base interactions, giving rise to a short-lived excited-state with emission that is slightly red-shifted compared with the 'normal' 2AP excited-state. Under these conditions, interaction of 2AP with any of the bases results in rapid non-radiative decay. The highly concentrated LiCl is

expected to dehydrate the dissolved dinucleotides, so that their local environment becomes hydrophobic. This would encourage the adoption of highly stacked conformations and enhance inter-base interaction. It is possible that aggregation of the dinucleotide into extended stacks might also occur, although this is unlikely at the low concentration used here. The change in the nature of the solvent environment is apparent in the value of τ_4 (and the lifetime of free 2APr) which decreases from ~10 ns in PBS to ~7 ns in LiCl, indicating that 2AP experiences a more hydrophobic environment in LiCl[21]. This dramatic effect of LiCl was not seen in the previous study of 2AP-containing duplexes[12], where there was little perturbation of the decay behaviour apart from a similar decrease in lifetime of extrahelical 2AP (τ_4), as shown in Table 5.1. Within the duplex the bases exist in an intrinsically hydrophobic environment and so inter-base interactions are little affected by the change in external solvent environment.

5.4.6 The Effect of Freezing

Freezing has a significant effect on the fluorescence decay of the 2AP-containing dinucleotides, causing the first two shortest decay components, with a total population more than 60% at room temperature, to vanish. As observed previously for DNA duplexes[12], freezing seems to eliminate the base motion in dinucleotides, and as a consequence block the fast ET quenching. τ_3 remains almost unchanged and A_3 decreases slightly at 77 K compared with room temperature, implying that this lifetime can be identified with the same conformation at room temperature and in the frozen matrix.

τ_4 at 77 K is ~14 ns, which is almost twice that at room temperature and approaching the predicted radiative lifetime of 2AP-A and 2AP-G dimer (~35 ns) by (TD)B3LYP method[2]. According to the calculated results, using the CIS method, obtained by Hardman *et al.*[3], there is delocalisation of electrons across 2AP and the neighbouring natural bases, and fluorescence transition is not entirely localised on 2AP. The long-lived excited-state seen in at 77 K may be due to a delocalised conformation

formed between 2AP and its neighbouring base, which is not subject to rapid ET quenching in the frozen matrix. This effect of freezing was not seen in the previous study of 2AP-labelled DNA duplexes[12], where τ_4 at 77 K is similar to that at room temperature, as shown in Table 5.1. Within the duplex the bases are in a more constrained environment due to the hydrogen bonding and a series of π-stacking interaction and therefore this type of conformation may not be accessible.

In summary, freezing of the dinucleotides eliminates charge transfer quenching. At 77 K, the dinucleotides exist in two static conformations: a weakly π-stacked conformation and a more strongly interacting, electron delocalised conformation.

5.5 Conclusion

In PBS buffer at room temperature 2AP-containing dinucleotides show detectable long-wavelength emission indicating that the 2AP heterodimer is formed in dinucleotides between 2AP and its 3' side base. However, the intensity of this emission is extremely weak suggesting that heterodimer formation is unfavourable. This supports the conclusion drawn in Chapter 4 that 2AP prefers to form a dimer with its 5' neighbour rather than with the base on the 3' side.

Overall, the fluorescence decay behaviour of 2AP in dinucleotides in PBS buffer at room temperature is similar to that in single strands and is described by 4 discrete components indicating comparable conformational heterogeneity in both cases. The shortest fluorescence lifetime of 2AP in dinucleotides varies with the identity of the neighbouring base and correlates with the driving force for ET, suggesting ET is the dominant mechanism of rapid quenching of 2AP* in dinucleotides.

The π-stacking interaction and the imperfectly π-stacked conformations in dinucleotides appears similar to that in single strands. However, the internal base motion and dynamic disorder in dinucleotides are greater than in single strands. The

population of highly stacked states decreases significantly and transfers to the weakly stacked states in dinucleotides compared with single strands. Thus, the series of base π-stacking interactions decreases the internal base motion and limits the conformational fluctuations of the single strand structure.

The fluorescence response of 2AP-containing dinucleotides is dramatically changed by dissolution in LiCl at room temperature, such that they all show a dominant, short lifetime of ~60 ps. The LiCl solvent appears to enhance the inter-base interactions, giving rise to a short-lived excited-state with emission that is slightly red-shifted compared with the 'normal' 2AP excited-state. The highly concentrated LiCl makes the dinucleotides local environment become hydrophobic, which seems encourage the adoption of highly stacked conformations and enhance inter-base interaction.

The fluorescence response of 2AP-containing dinucleotides is significantly affected by freezing to 77 K, such that the first two shortest decay components vanish. Freezing eliminates the base motion, and as a consequence blocks the fast ET quenching. The frozen dinucleotides show a delocalised conformation formed between 2AP and its neighbouring base with an increased lifetime ~14 ns, which is significantly longer than that of the free 2APr under the same conditions. The majority of the conformational population, in which 2AP was highly quenched at room temperature, exists in the unquenched delocalized conformation at 77 K.

5.6 References

(1) Kelley, S. O.; Barton, J. K. *Science* **1999**, *283*, 375-381.

(2) Jean, J. M.; Hall, K. B. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 37-41.

(3) Hardman, S. J. O.; Thompson, K. C. *Int. J. Quantum Chem.* **2007**, *107*, 2092-2099.

(4) Somsen, O. J. G.; van Hoek, A.; van Amerongen, H. *Chem. Phys. Lett.* **2005**, *402*, 61-65.

(5) Somsen, O. J. G.; Keukens, L. B.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H. *ChemPhysChem* **2005**, *6*, 1622-1627.

(6) O'Neill, M. A.; Becker, H; Wan, C.; Barton, J. K.; Zewail, A. H. *Angew. Chem. Int. Edit.* **2003**, *42*, 5896-5900.

(7) Somsen, O. J. G.; Trinkunas, G.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H. *J. Lumin.* **2006**, *119-120*, 100-104.

(8) Shlesinger, M. F.; Montroll, E. W. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 1280-1283.

(9) Gidden, J.; Bowers, M. T. *Eur. Phys. J. D* **2002**, *20*, 409-419.

(10) Narayanan, M.; Kodali, G.; Xing, Y.; Stanley, R. J. *J. Phys. Chem. B* **2010**, *114*, 10573-10580.

(11) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Springer; 3rd edition: 2006.

(12) Neely, R. K.; Jones, A. C. *J. Am. Chem. Soc.* **2006**, *128*, 15952-15953.

(13) O'Neill, M. A.; Barton, J. K. *J. Am. Chem. Soc.* **2004**, *126*, 13234-13235.

(14) Cai, Z.; Li, X.; Sevilla, M. D. *J. Phys. Chem. B* **2002**, *106*, 2755-2762.

(15) Bonnist, E. Y. M.; Jones, A. C. *ChemPhysChem* **2008**, *9*, 1121-1129.

(16) Bonnist, E. Y. M., PhD thesis, The Univeristy of Edinburgh, **2008**.

(17) Wan, C.; Xia, T.; Becker, H.; Zewail, A. H. *Chem. Phys. Lett.* **2005**, *412*, 158-163.

(18) Steenken, S.; Jovanovic, S. *J. Am. Chem. Soc.* **1997**, *119*, 617-618.

(19) Larsen, O. F. A.; van Stokkum, I. H. M.; de Weerd, F. L.; Vengris, M.; Aravindakumar, C. T.; van Grondelle, R.; Geacintov, N. E.; van Amerongen, H. *Phys. Chem. Chem. Phys.* **2004**, *6*, 154-160.

(20) Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G. *J. Phys. Chem. B* **2007**, *111*, 5757-5766.

(21) Neely, R. K., PhD thesis, The University of Edinburgh, **2005**.

Chapter 6

2-Aminopurine as a Probe of DNA-Protein Interactions: the Pfu-Pol Archaeal Family-B DNA Polymerase

6.1 Introduction

DNA polymerase is an enzyme, the function of which is to synthesise DNA. It plays an important role in DNA replication and repair. It enables the replication of DNA in the 5'-3' direction of the primer and is a template-directed enzyme, which means that the base added to the primer strand is complementary to the base on the template strand. Polymerisation of DNA only occurs when the incoming base forms a Watson-Crick base pair with the base on the template strand. However, some base analogues can be added on the primer strand by the DNA polymerase by mistake, if they can form Watson-Crick type base pairs with the template strand. Examples of such base analogues are uracil (U), formed by deamination of cytosine (C), is a base analogue of thymine (T) and forms a Watson-Crick base pair with adenine (A). Hypoxanthine (H), which is deaminated adenine (A), is an analogue of guanine (G) and can base pair with cytosine (C). If an error is made in DNA replication, DNA polymerase will hydrolyse the DNA from the 3'-hydroxyl terminus of the primer strand and remove the misincorporated mononucleotide. Therefore, DNA polymerase also contains a 3'-5' proofreading exonuclease. The polymerase always checks the result of each base added before catalysing addition of the next base to the primer strand. This editing function greatly reduces the error because base pairs are examined twice by the polymerase. DNA replication is extremely precise. The error rate is 1 out of 10^8 to 10^{10} base pairs copied.

Depending on the amino acid sequence similarities, DNA polymerases are normally classified into four major groups: the family A, B, C and X[1]. Family-B DNA polymerases are the main replicative DNA polymerases from eucaryotic cells. *Bacteria*, *Eucarya* and *Archaea* are the three major branches of life[2]. All of the archaeal DNA polymerases are classified as family-B DNA polymerases. Archaeal DNA polymerases have a remarkable property that they bind tightly to the deaminated base U or H and stop DNA replication when these bases are encountered[3-10]. A notable feature of the archaeal family-B polymerase is that it recognises the deaminated base in the single stranded region of the DNA template, whereas other polymerases recognise the mismatched base pair at the terminus of the 3' primer.

The polymerase and exonuclease domains of DNA polymerases are always well separated and unwinding of the primer to expose a short single stranded region is required to position the 3′-terminal base of the primer in the exonuclease active site[11-15]. Connolly and coworkers[4] proposed that if the polymerase approaches nearer than four bases to U, the primer starts to unravel, giving a short single stranded region at its 3′ end, as shown in Figure 6.1. Such separation explains why primer-templates with U at +1, +2, +3 are still capable of interacting strongly with the polymerase. A more detailed description of archaeal family-B DNA polymerase and its function is given in Appendix B.

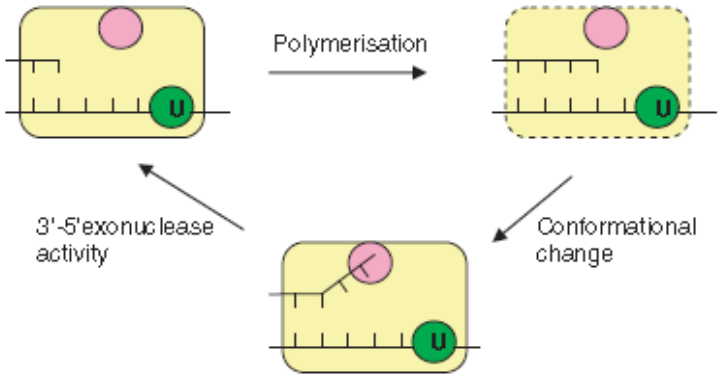


Figure 6.1 Idling by Pfu-Pol on encountering a template strand deaminated bases. Reproduced from reference 4. The polymerase, which is yellow, captures U at +4 posision from the primer-template junction using a specific binding pocket, which is green. The 3'-5' exonuclease site is rose.

In this chapter, the hypothesis that the proof-reading mechanism of archeal family-B polymerase involves unwinding the primer-template duplex region, upon encountering the misincorpated U, is investigated by time-resolved fluorescence measurements. Time-resolved fluorescence of 2-aminopurine (2AP) is used here for the first time to investigate the role played by unwinding of the primer-template in this mechanism. As shown previously in Chapter 4, the fluorescence decay of 2AP can successfully report the effect of unpairing on the conformation and hence on inter-base interaction in DNA. Here, fluorescence decays were measured for 2AP-labelled primer-templates, with a misincorpated U at the +2 position (U+2) or +4 position (U+4) from the replication fork, and the 2AP-labelled primer single strand, in complex with the wild type polymerase Pfu-Pol (WT) or a mutant of Pfu-Pol without exonuclease activity (exo$^-$). The sequences of the 2AP-labelled primer-templates and the 2AP-labelled primer single strand are listed in Table 6.1.

Oligodeoxynucleotide (abbreviation)	Oligodeoxynucleotide sequence
U+2 Primer-template (U+2)	5'-(Cy5)-GGGGATCCTCTAGAGTCGACCTGCAGGGC**P**A-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTC**U**TTCGAACAGAGG-5'
U+4 Primer-template (U+4)	5'-(Cy5)-GGGGATCCTCTAGAGTCGACCTGCAGGGC**P**A-3' 3'-CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCTT**U**CGAACAGAGG-5'
Primer single strand (ss)	5'-(Cy5)-GGGGATCCTCTAGAGTCGACCTGCAGGGC**P**A-3'

Table 6.1 Base sequences of 2AP-labelled primer-templates and 2AP-labelled primer single strand. P denotes 2AP and U stands for uracil. Duplexes are named by the U position from the replication fork.

6.2 Materials and Methods

2AP-labelled oligonucleotides and archaeal family-B DNA polymerases were gifts from Prof. Bernard A. Connolly, the University of Newcastle. The sequences of 2AP-labelled primer-templates and 2AP-labelled primer single strand are shown in Table 6.1. The primer-templates were buffered in 10mM Hepes-NaOH (pH 7.5), 100 mM NaCl, 1 mM EDTA, whereas the primer single strand was stored in distilled water. The concentration of primer-templates and primer single strand used was 2

μM. Two polymerases: wild type Pfu-Pol with exonuclease activity (WT) and a mutant without exonuclease activity (exo$^-$) were used. The storage buffer for both polymerases was 300 mM NaCl, 25 mM Tris-HCl (pH 8.0) and 1 mM EDTA. The concentration of the polymerases used was 8 μM in order to ensure complete binding to 2AP-labelled primer-templates. Reactions were carried out in two different buffers: Ca^{2+} buffer containing 20 mM Tris-HCl (pH 8.5), 10 mM KCl, 20 mM $CaCl_2$; and Mg^{2+} buffer including 20 mM Tris-HCl (pH 8.5), 10 mM KCl, 20 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$.

Time-resolved fluorescence was recorded using the TCSPC technique, as described in Chapter 3, at an excitation wavelength 315 nm. The decay curves were normally collected at three emission wavelengths: 365 nm, 380 nm and 395 nm, with the stop condition for collection that the peak of the decay curve had reached 10,000 counts. The fluorescence decays of 2AP riboside (2APr) were collected under the same condition as primer-templates but at a single emission wavelength of 380 nm. To investigate the reaction process of the exonuclease-active WT enzyme, the fluorescence decays of 2AP in the primer-template-WT complexes were recorded at a single emission wavelength of 380 nm, with each decay being collected for 1 minute. The slits of the monochromator were set to 14 nm bandpass.

6.3 Results and Discussion

6.3.1 2AP-labelled DNA

The fluorescence decay parameters of the two primer-templates and one primer single strand in Ca^{2+} and Mg^{2+} buffers at excitation wavelength 315 nm are presented in Table 6.2. The fluorescence lifetimes of 2APr in the corresponding buffers are also included for comparison. Amplitude parameters (A) are for emission wavelength 380 nm. The full set of the decay parameters for the two primer-templates and primer single strand in Ca^{2+} and Mg^{2+} buffers at all three emission wavelengths are given in Tables A36 to A41 in Appendix A. Previous studies[16,17] have shown that Ca^{2+} does

not support proof reading exonucleolysis, whereas Mg^{2+} does. Therefore, Ca^{2+} and Mg^{2+} buffers were used to investigate whether the different metal ions affect the DNA conformational changes or the polymerase Pfu-Pol activity.

Sample	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
U+2/Ca2+	0.04	0.64	2.8	8.1	0.74	0.14	0.09	0.03
U+2/Mg2+	0.05	0.63	2.7	8.0	0.71	0.15	0.10	0.04
U+4/Ca2+	0.04	0.46	2.5	8.2	0.86	0.08	0.04	0.02
U+4/Mg2+	0.04	0.44	2.4	8.2	0.86	0.08	0.04	0.02
ss/Ca2+	0.05	0.55	2.4	6.9	0.68	0.16	0.11	0.05
ss/Mg2+	0.05	0.54	2.3	6.7	0.68	0.17	0.11	0.04
2APr/Ca2+	-	-	-	9.9	-	-	-	1
2APr/Mg2+	-	-	-	9.8	-	-	-	1

Table 6.2 The global fluorescence decay parameters for the two primer-templates and primer single strand in Ca^{2+} and Mg^{2+} buffers excited at 315 nm. A factors are for 380 nm emission. The fluorescence lifetimes of 2APr are also shown. Experimental uncertainties (estimated by dividing standard deviation of values by the corresponding values): τ_1 (≤ 5%), τ_2 (≤ 3%), τ_3 (≤ 2%), τ_4 (≤ 1%), A_1 (≤ 1%), A_2 and A_3 (≤ 2%), and A_4 (≤ 3%).

The fluorescence decay behaviours of all the 2AP-labelled primer-templates and primer single strand in Ca^{2+} buffer are almost identical to the corresponding ones in Mg^{2+} buffer, suggesting that in all cases the conformational behaviour is independent of the identity of the metal ion. Therefore, the following discussion uses the decay parameters in Ca^{2+} buffer as an example.

It can be seen that the fluorescence decays of the two 2AP-labelled primer-templates and primer single strand are multi-exponential characterised by four discrete components. The fluorescence decay behaviour of 2AP in the two primer-templates is consistent with that of 2AP-labelled duplexes (as described in Section 2.2.2.3 of Chapter 2). The fluorescence decays of 2AP in U+2 and U+4 are dominated (≥ 74%) by a short component of ~50 ps. Less than 3% of the 2AP molecules have fluorescence lifetimes around 8 ns. The two intermediate lifetimes in U+2 and U+4 are ~0.6 ns and ~2.5 ns. The A-factors decrease with increasing the fluorescence lifetimes. There is almost no dependence of the amplitudes of lifetimes on different emission wavelength (see Tables A36 to A39 in Appendix A), implying that the

decay components have very similar emission spectra. In general, the fluorescence decay behaviour of 2AP in U+2 is very similar to that in U+4 except the highly stacked conformation of U+4 is even more populated (86%).

The fluorescence decay behaviour of the 2AP-labelled primer single strand is also consistent with that of other 2AP-labelled single strands (as described in Chapter 4). The value of A_1, 0.68, is decreased slightly compared with U+2 and U+4 and the population is transferred to the other three components. The magnitude of τ_4 of primer single strand is decreased slightly compared with those in U+2 and U+4, due to the effect of increasing internal base motion in the single strand, as discussed in Chapter 4. Broadly speaking, the change of 2AP fluorescence decay behaviour from U+2 and U+4 to primer single strand is consistent with the change from DNA duplex to single strand seen previously, as discussed in Chapter 4.

6.3.2 2AP-labelled Primer-templates and exo$^-$ Complexes

Table 6.3 shows the fluorescence decay parameters of the 2AP-labelled primer-templates upon binding to a mutant of polymerase Pfu-Pol (exo$^-$), which does not have the exonuclease activity. Amplitude parameters (A) are for emission wavelength 380 nm. The full set of the decay parameters for the U+2-exo$^-$ and U+4-exo$^-$ complexes in Ca^{2+} and Mg^{2+} buffers at all three emission wavelengths are given in Tables A42 to A45 in Appendix A.

Sample	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
U+2-exo-/Ca2+	0.17	0.94	3.6	9.0	0.40	0.30	0.20	0.10
U+2-exo-/Mg2+	0.17	0.99	3.7	8.7	0.35	0.25	0.25	0.15
U+4-exo-/Ca2+	0.06	0.67	3.6	9.5	0.60	0.11	0.17	0.12
U+4-exo-/Mg2+	0.06	0.67	3.5	9.4	0.60	0.12	0.17	0.11

Table 6.3 The global fluorescence decay parameters for the U+2-exo$^-$ and U+4-exo$^-$ complexes in Ca^{2+} and Mg^{2+} buffers excited at 315 nm. A factors are for 380 nm emission. Experimental uncertainties (estimated by dividing standard deviation of values by the corresponding values): τ_1 ($\leq$ 5%), τ_2 ($\leq$ 4%), τ_3 ($\leq$ 2%), τ_4 ($\leq$ 1%), A_1 ($\leq$ 1%), and A_2, A_3 and A_4 ($\leq$ 2%).

The fluorescence decay behaviour of the U+2-exo$^-$ and U+4-exo$^-$ complexes in Ca^{2+} buffer is very similar to the corresponding ones in Mg^{2+} buffer, implying different metals have little effect on the enzyme binding. Therefore, the following discussion uses the decay parameters in Ca^{2+} buffer as an example.

6.3.2.1 U+2-exo$^-$ Complex

A plot of the decay parameters of 2AP in U+2 and U+2-exo$^-$ complex is shown in Figure 6.2. There is a substantial change to decay parameters upon exo$^-$ binding. The magnitudes of all the fluorescence lifetimes of 2AP increase in U+2-exo$^-$ complex compared with those in U+2, indicating the loss of quenching and change in local environment upon exo$^-$ binding. In particular, the magnitude of shortest lifetime component in U+2-exo$^-$ complex increases to 170 ps that is more than four times as big as that in U+2, implying a decrease in the extent of π-stacking interaction. The amplitude of the shortest lifetime component decreases dramatically from 0.74 in U+2 to 0.40 in U+2-exo$^-$ with an increase in the each of the other three components, A_2 to A_4, indicating a large transfer of population from highly stacked to poorly stacked conformations.

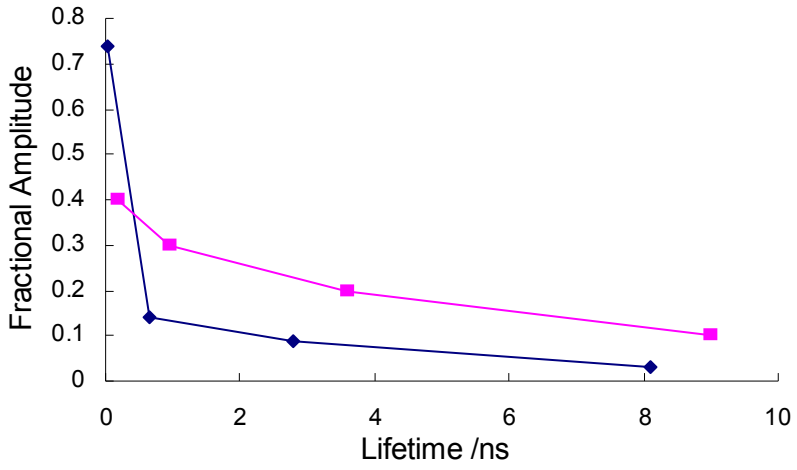


Figure 6.2 A plot of the decay parameters of 2AP in U+2 (blue line) and U+2-exo$^-$ complex (purple line) in Ca^{2+} buffer.

The change in decay parameters on enzyme binding is consistent with the change in 2AP fluorescence intensity from steady-state measurement performed by Richardson

et al.[18] on the same samples. The fluorescence decay behaviour of 2AP in U+2-exo$^-$ complex is entirely consistent with unwinding of the duplex, leaving 2AP in a poorly stacked, unpaired state. This observation is consistent with the crystal structure of the H+2-Tgo-Pol (one type of wild type archeal family-B polymerase) complex with a primer-template mimic[19], which shows that the last two bases from 3' side of the primer become single strand when bound to the polymerase, as seen in Figure B3 (2) in Appendix B. The crystal structure[19] shows that the stacking interaction is interrupted and primer strand unwinds from the complementary bases in the template strand, which generates a two base long single stranded primer extension.

6.3.2.2 U+4-exo$^-$ Complex

A plot of the decay parameters of 2AP in U+4 and the U+4-exo$^-$ complex is shown in Figure 6.3. The shortest lifetime component in the U+4-exo$^-$ complex remains very short, 60 ps, only slightly increased compared with that in U+4 (40 ps), indicating the highly stacked conformation that facilitates the fast electron transfer quenching is almost unchanged upon exo$^-$ binding. The population of this conformation decrease from ~90% to 60% upon exo$^-$ binding, but remains predominant. The magnitude of the second component is somewhat increased on enzyme binding but remains substantially shorter than in the U+2-exo$^-$ complex. The third lifetime also increases and is comparable to that in the U+2-exo$^-$ complex. A_2 increase only slightly in the exo$^-$ complex, whereas A_3 and A_4 increase more than 10%. There is thus a transfer of population from highly stacked to poorly stacked states. Again, the decay parameters are consistent with the steady state measurements of Richardson *et al*.[18].

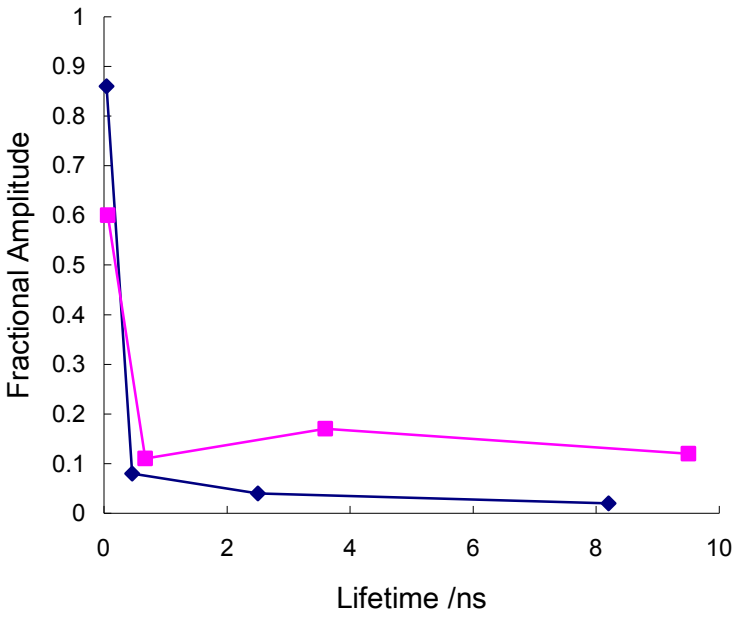


Figure 6.3 A plot of the decay parameters of 2AP in U+4 (blue line) and U+4-exo$^-$ complex (purple line) in Ca^{2+} buffer.

Overall, the decay parameters for U+4 bound to exo$^-$ indicates a modest distortion of the duplex in the region of the primer-template junction. Again, this is consistent with the crystal structure of U+4-Tgo-Pol complex[7], which indicates that there is no interaction between 3' side terminus bases in the primer and the polymerase binding active site, and no major structural changes in the duplex region of the primer-template are observed.

6.3.2.3 Comparison of U+2-exo$^-$ and U+4-exo$^-$ Complexes

Although, deaminated bases at +4 position and +2 position from the replication fork of primer-templates binding to the polymerase are recognised by the exactly the same pocket[7,19], the fluorescence decays of 2AP in U+2-exo$^-$ and U+4-exo$^-$ complexes are significantly different. In particular, the shortest lifetime component of 2AP in U+4-exo$^-$ complex (60 ps) is roughly one-third as much as that in U+2-exo$^-$ complex, and its population (60%) is one and half times that in U+2 complex, indicating the continued existence of a highly stacked state in U+4 complex. The value of τ_2 in U+4-exo$^-$ complex also remains significantly shorter than in U+2-exo$^-$ complex, again indicating the persistence of better-stacked conformations. Therefore, binding

of exo$^-$ has much less effect on the conformation of U+4 than U+2. It is also consistent with the unwinding of primer-templates mechanism proposed by Connolly group[4].

6.3.3 2AP-labelled Primer-templates and WT Complexes

In this section, binding of 2AP-labelled primer-templates upon to the wild type polymerase Pfu-Pol (WT), which has exonuclease activity, is considered. The decay parameters are presented in Table 6.4. 2AP in both U+2-WT and U+4-WT complexes, in both buffers, has only one lifetime component of ~7.2 ns, indicating that 2AP experiences a homogeneous environment. Therefore, at least the last two bases from the 3' terminus of primer in the duplex region has been fully cut by WT. However, the 7.2 ns lifetime is considerably shorter than that that in water solution (10.6 ns)[20]. According to Neely *et al.*[20], the fluorescence lifetime of 2AP decreases with decreasing solvent polarity, thus the cut 2AP inside the WT complexes experiences a less polar environment than in water solution. Although Ca^{2+} does not support proof reading exonucleolysis[16,17], it does not inhibit the exonuclease activity.

Sample	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
U+2-WT/Ca^{2+}	-	-	-	7.2	-	-	-	1
U+2-WT/Mg^{2+}	-	-	-	7.3	-	-	-	1
U+4-WT/Ca^{2+}	-	-	-	7.2	-	-	-	1
U+4-WT/Mg^{2+}	-	-	-	7.2	-	-	-	1

Table 6.4 The global fluorescence decay parameters for the U+2-exo$^-$ and U+4-exo$^-$ complexes in Ca^{2+} and Mg^{2+} buffers excited at 315 nm. Experimental uncertainties (estimated by dividing standard deviation of values by the corresponding values):τ_4 and A_4 (< 1%).

6.3.4 Reaction Process in 2AP-labelled Primer-templates and WT Complexes

From above, WT polymerase can cut the 3' terminus bases of primer in both Ca^{2+} and Mg^{2+} buffers. In order to investigate whether the progress of the reaction could be observed by the change in 2AP fluorescence decay, the decays were recorded in 1 minute intervals. The first measurement (0 minute) was made about 30 seconds initial mixing of enzyme and substrate. The quality of the decay recorded for only 1 minute is relatively low, which prevents a detailed analysis and interpretation of the results.

6.3.4.1 U+2-WT Complex

Fluorescence decay parameters for the U+2-WT complex in Ca^{2+} buffer at excitation wavelength 315 nm are shown in Table 6.5, along with chi-square values for the fits. 2AP in U+2-WT complex in Ca^{2+} buffer has two fluorescence lifetimes after 1 minute, indicating the second base (2AP) in the 3' side of primer is not fully excised by the exonuclease of WT. After 3 minutes the 2AP is cut and removed by the exonuclease of WT as shown by the existence of only one fluorescence lifetime component.

Time/ minute	2 components			1 component		
	τ_i/ ns	A_i	χ^2	τ_i/ ns	A_i	χ^2
1	1.1 7.3	0.36 0.64	1.017	6.6	1	1.829
2	1.3 7.2	0.10 0.90	1.013	7.0	1	1.096
3	1.3 7.2	0.05 0.95	1.021	7.1	1	1.043
4	6.8 7.9	0.22 0.78	1.033	7.1	1	1.033

Table 6.5 The fluorescence decay parameters for the U+2-WT complex in Ca^{2+} buffer excited at 315 nm, as a function of reaction time, commencing 30 seconds after mixing. Local chi-squares for the fits are also shown.

Fluorescence decay parameters for the U+2-WT polymerase complex in Mg^{2+} buffer are shown in Table 6.6, along with chi-square values for the fits. The 2AP in U+2-WT complex in Mg^{2+} buffer only has one component after one minute, indicating the second base (2AP) is fully excised by the exonuclease of WT after 1 minute.

Time/ minute	2 components			1 component		
	τ_i/ ns	A_i	χ^2	τ_i/ ns	A_i	χ^2
1	1.0 7.4	-0.006 1.006	1.126	7.4	1	1.122
2	0.17 7.3	0.002 0.998	1.075	7.3	1	1.072

Table 6.6 The fluorescence decay parameters for the U+2-WT complex in Mg^{2+} buffer excited at 315 nm, as a function of reaction time, commencing 30 seconds after mixing. Local chi-squares for the fits are also shown.

Therefore, the exonuclease of WT works more efficient in Mg^{2+} buffer than in Ca^{2+} buffer, which is consistent with previously studies[16,17] that metal Ca^{2+} does not support proof reading exonucleolysis, whereas Mg^{2+} does.

6.3.4.2 U+4-WT Complex

Fluorescence decay parameters for the U+4-WT complex in Ca^{2+} buffer at excitation wavelength 315 nm are shown in Table 6.7, along with chi-square values for the fits. 2AP in U+4-WT complex in Ca^{2+} buffer has at least two, and possibly three, resolvable lifetime components after 1 minute. The excision of 2AP in U+4 is complete after 7 minutes, indicating process of excising bases from the 3' terminus is slower than that in U+2 in the same buffer. This is consistent with the observation by Russell *et al.*[4], from exonuclease assay gel images, that the primer-template of U+2 in Mg^{2+} buffer is broken down by WT very quickly with the majority of the initial primer excised in 10 seconds, whereas most of the primer of U+4 persists after 2 minutes. Although, the Mg^{2+} buffer was used in Russell *et al.*[4] assay, their result show a similar reaction rate difference between U+2 and U+4.

Time/ minute	3 components			2 components			1 component		
	τ_i/ ns	A_i	χ^2	τ_i/ ns	A_i	χ^2	τ_i/ ns	A_i	χ^2
1	0.23 1.7 6.6	0.27 0.33 0.40	1.070	- 1.0 6.2	- 0.48 0.52	1.110	- - -	- - -	- - -
2	0.46 2.0 7.0	0.19 0.26 0.55	1.013	- 1.2 6.8	- 0.37 0.63	1.024	- - -	- - -	- - -
3	- -	- -	- -	1.6 7.1	0.25 0.75	1.044	6.0	1	1.233
4	- -	- -	- -	1.4 7.1	0.14 0.86	1.033	6.8	1	1.156
5	- -	- -	- -	1.5 7.1	0.07 0.93	1.005	7.0	1	1.040
6	- -	- -	- -	1.3 7.1	0.06 0.94	1.005	7.0	1	1.044
7	- -	- -	- -	1.8 7.1	0.05 0.95	0.958	7.0	1	0.977

Table 6.7 The fluorescence decay parameters for the U+4-WT complex in Ca^{2+} buffer excited at 315 nm, as a function of reaction time, commencing 30 seconds after mixing. Local chi-squares for the fits are also shown.

Fluorescence decay parameters for the U+4-WT complex in Mg^{2+} buffer at excitation wavelength 315 nm are shown in Table 6.8, along with chi-square values for the fits. The 2AP decay has only one component after one minute, indicating that 2AP is removed by the exonuclease of WT after 1 minute. This confirms the assertion above that the efficiency of exonuclease on excision bases is higher in Mg^{2+} buffer than in Ca^{2+} buffer and is again consistent with the report of Russell *et al.*[4].

Time/ minute	2 components			1 component		
	τ_i/ ns	A_i	χ^2	τ_i/ ns	A_i	χ^2
1	0.08 7.4	-0.1 1.1	1.025	7.4	1.0	1.035

Table 6.8 The fluorescence decay parameters for the U+4-WT complex in Mg^{2+} buffer excited at 315 nm, as a function of reaction time, commencing 30 seconds after mixing. Local chi-squares for the fits are also shown.

6.4 Conclusion

The 2AP fluorescence decay has been used successfully to investigate the structural distortion of primer-templates by an archaeal family-B DNA polymerase which lacks exonuclease activity. The fluorescence decay of 2AP in the U+2-exo$^-$ complex shows a dramatical increase in the lifetime of the shortest decay component, representing a less highly stacked 2AP conformation, and provides convincing evidence for a decrease in the amount of double stranded DNA on polymerase exo$^-$ binding. In contrast, a much smaller perturbation of the 2AP fluorescence decay is seen in the U+4-exo$^-$ complex, indicating less profound DNA distortion. It is concluded that the presence of U at the +2 position leads to more DNA unwinding than observed with U+4, but even with U+4 some unravelling may occur. The structural distortions inferred here are consistent with the crystal structural studies of the primer-templates-polymerase complex, and support the hypothesis that the mechanism of archeal family-B polymerase involves unwinding of the primer-template duplex region upon encountering the misincorporated U.

2AP fluorescence also successfully reported the reaction process of the exonuclease of the wild type polymerase, excising the 3' terminus primer bases in primer-templates, albeit with limited time resolution. The speed of excising the first two bases from 3' terminus of primer in U+2 is much faster than in U+4. In addition, the speed of excising bases in the 3' terminus of primer in Mg^{2+} buffer is significantly faster than in Ca^{2+} buffer. These results are consistent with those from a previous polymerase excision gel assay.

6.5 References

(1) Ito, J.; Braithwaite, D. K. *Nucleic Acids Res.* **1991**, *19*, 4045-4057.

(2) Boulet, A.; Simon, M.; Faye, G.; Bauer, G. A.; Burgers, P. M. J. *EMBO J.* **1989**, *8*, 1989-1954.

(3) Greagg, M. A.; Fogg, M. J.; Panayotou, G.; Evans, S. J.; Connolly, B. A.; Pearl, L. H. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9045-9050.

(4) Russell, H. J.; Richardson, T. T.; Emptage, K.; Connolly, B. A. *Nucleic Acids Res.* **2009**, *37*, 7603-7611.

(5) Fogg, M. J.; Pearl, L. H.; Connolly, B. A. *Nat. Struct. Biol.* **2002**, *9*, 922-927.

(6) Shuttleworth, G.; Fogg, M. J.; Kurpiewski, M. R.; Jen-Jacobson, L.; Connolly, B. A. *J. Mol. Biol.* **2004**, *337*, 621-634.

(7) Firbank, S. J.; Wardle, J.; Heslop, P.; Lewis, R. J.; Connolly, B. A. *J. Mol. Biol.* **2008**, *381*, 529-539.

(8) Connolly, B. A. *Biochem. Soc. Trans.* **2009**, *37*, 65-68.

(9) Gill, S.; O'Neill, R.; Lewis, R. J.; Connolly, B. A. *J. Mol. Biol.* **2007**, *372*, 855-863.

(10) Gruz, P.; Shimizu, M.; Pisani, F. M.; Felice, M. D.; Kanke, Y.; Nohmi, T. *Nucleic Acids Res.* **2003**, *31*, 4024-4030.

(11) Joyce, C. M. *J. Biol. Chem.* **1989**, *264*, 10858-10866.

(12) Joyce, C. M.; Steitz, T. A. *Annu. Rev. Biochem.* **1994**, *63*, 777-822.

(13) Brautigam, C. A.; Steitz, T. A. *Curr. Opin. Struc. Biol.* **1998**, *8*, 54-63.

(14) Shamoo, Y.; Steitz, T. A. *Cell* **1999**, *99*, 155-166.

(15) Reha-Krantz, L. J. *Biochim. Biophys. Acta.* **2010**, *1804*, 1049-1063.

(16) Beese, L. S.; Steitz, T. A. *EMBO J.* **1991**, *10*, 25-33.

(17) Datta, K.; Johnson, N. P.; LiCata, V. J.; Von Hippel, P. H. *J. Biol. Chem.* **2009**, *284*, 17180-17193.

(18) Richardson, T. T.; Wu, X.; Dryden, D. T. F.; Jones, A. C.; Connolly, B. A. *Paper is under preparation*.

(19) Killelea, T.; Ghosh, S.; Tan, S. S.; Heslop, P.; Firbank, S. J.; Kool, E. T.; Connolly, B. A. *Biochemistry* **2010**, *49*, 5772-5781.

(20) Neely, R. K., PhD thesis, The University of Edinburgh, **2005**.

Chapter 7

Experimental and Computational Studies on Selenium Base Analogues

7.1 Introduction

The work reported in Chapters 4, 5 and 6 used 2-aminopurine (2AP) as a fluorescent probe to investigate the conformational changes, interactions and dynamics of DNA. In recent years, a new class of base analogues in which selenium (Se) replaces oxygen (O) at the 4-position in thymine (^{Se}T) or at 6-position in guanine (^{Se}G) bases have been synthesised. These new base analogues are of potential interest as spectroscopic probes because they show absorption peaks that are greatly red-shifted (by ~100 nm) compared with the natural bases[1,2], taking their absorption spectra into the visible region. In order to examine whether ^{Se}T and ^{Se}G can be used as fluorescent base analogues, the UV absorption spectra, fluorescence and phosphorescence emission spectra of 4-Se-thymidine-3'-phosphate (^{Se}T*p) and 6-Se-2'-deoxyguanosine-3'-phosphate (^{Se}G*p) have been studied, as reported in this chapter. The structures of ^{Se}T*p and ^{Se}G*p are shown in Figure 7.1. To investigate structural difference between the Se-base analogues and the corresponding natural bases, and the origins of the different photophysical properties, the base analogues, ^{Se}T, and ^{Se}G, and the corresponding natural bases, T and G, were studied by quantum chemical calculations. The structures of these molecules are shown in Figure 7.2.

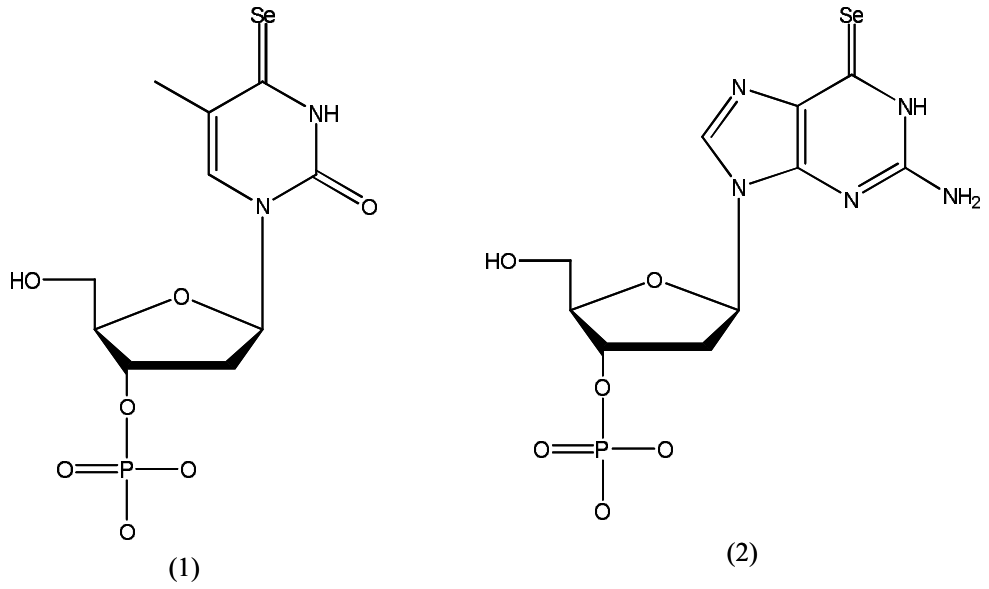


Figure 7.1 The structures of (1) ^{Se}T*p and (2) ^{Se}G*p.

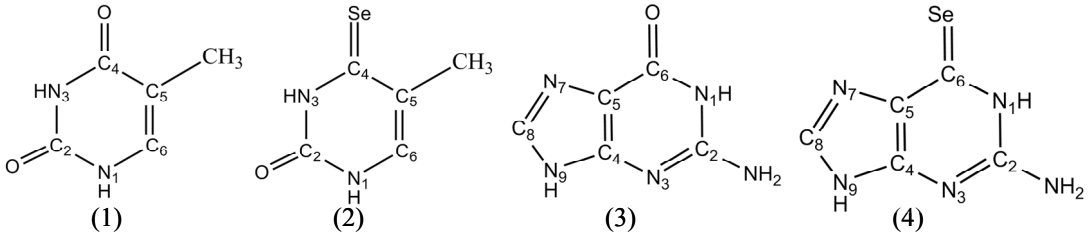
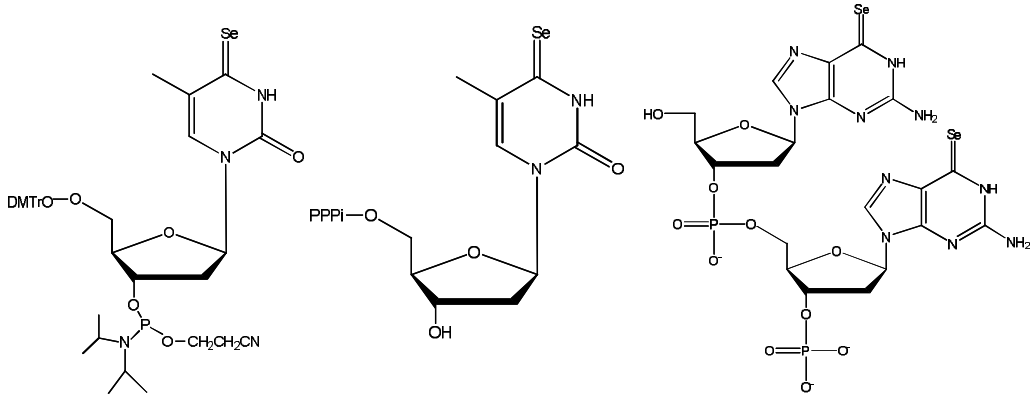


Figure 7.2 Molecular structures and atom numbering of (1) T, (2) ^{Se}T, (3) G and (4) ^{Se}G.

4-Se-thymidine phosphoramidite (^{Se}TPP), shown in Figure 7.3 (1), in which O at the 4-position of thymidine has been replaced with a Se atom, was first synthesised by Salon *et al.*[3]. The ^{Se}TPP has been incorporated into oligonucleotides to investigate the ^{Se}T - adenine (A) base pair formation and selectivity, the duplex thermostability and recognition, and the replication efficiency and fidelity[3]. A stability study of 5'-ATGG^{Se}TGCTC-3' showed that ^{Se}T was stable at 60 °C over 3 hours without significant decomposition[3]. The UV-melting temperature of the ^{Se}T-containing DNA duplex was determined to be 38.6 °C, which is similar to that of the corresponding normal DNA duplex (39.2 °C). Salon *et al.*[3] also obtained the crystal structure of the ^{Se}T-containing DNA duplex mentioned above. Structural analysis indicated that replacement of O by Se did not perturb the structure significantly and the crystal structure of the ^{Se}T-containing DNA duplex was almost identical to that of the

corresponding native DNA. The DNA duplex is flexible enough to accommodate ^{Se}T with a slight shift of Se. The crystal structure also reveals that ^{Se}T forms a Se-mediated hydrogen bond with the 6-amino group of adenosine within the ^{Se}T-A base pair.



(1) (2) (3)

Figure 7.3 The structures of (1) ^{Se}TPP, (2) ^{Se}TTP and (3) ^{Se}G^{Se}G dimer.

Later, Caton-Williams *et al.*[1] successfully synthesized 4-Se-thymidine 5'-triphosphate (^{Se}TTP), in which O at the 4-position of thymidine was replaced by a Se atom, as shown in Figure 7.3 (2). They found the aqueous solution of ^{Se}TTP to be yellow, whereas native thymidine 5'-triphosphate (TTP) is colourless in aqueous solution. The UV absorption spectrum of ^{Se}TTP has a peak at 369 nm (extinction coefficient $\varepsilon = 1.8 \times 10^4$ M^{-1} cm^{-1}) and is red-shifted by about 100 nm relative to that of TTP, which has its main absorption band at 267 nm ($\varepsilon = 9.2 \times 10^3$ M^{-1} cm^{-1}), as shown in Figure 7.4. The extinction coefficient of the 369 nm peak of ^{Se}TTP is approximately twice that of the 267 nm peak of TTP. Besides that, there is another peak at about 205 nm for both TTP and ^{Se}TTP. Caton-Williams *et al.*[1] attributed the observed red-shifted absorption peak to the ease of delocalization of the Se electrons on the nucleobase. However, by examining the UV absorption spectrum of ^{Se}TTP carefully, there is evidence of a weak absorption at about 280 nm that is similar to the peak of TTP. Therefore, the 369 nm absorption peak of ^{Se}TTP may result from the same transition as that of the 267 nm peak of TTP as Caton-Williams *et al.* suggested or it may be a completely new transition. Electronic structure calculations on the

corresponding molecules performed in the later part of this chapter address this question.

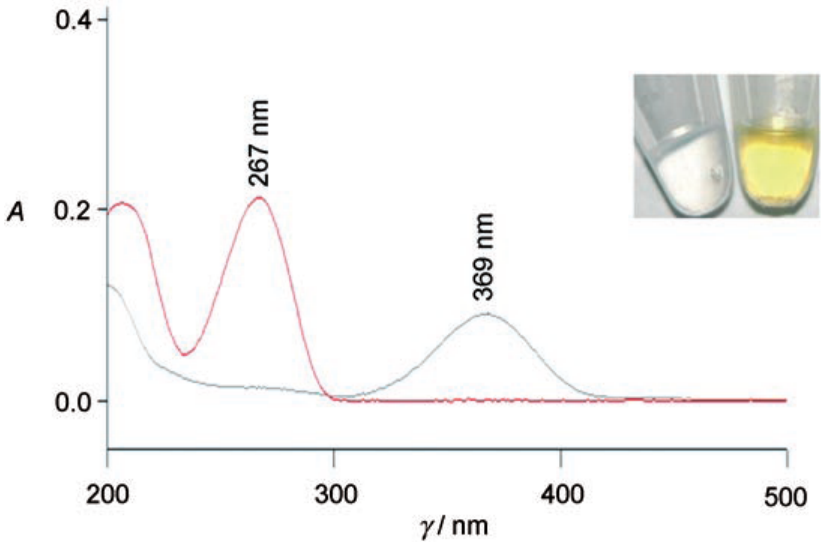


Figure 7.4 The UV absorption spectra of TTP (red line) and ^{Se}TTP (grey line). The small figure in the right corner: TTP (left, coulorless) and ^{Se}TTP (right, yellow). Reproduced from reference 1.

Salon *et al.*[2] synthesised, purified and measured the absorption spectra of GG, G^{Se}G (in which the 6-position O has been replaced by Se) and ^{Se}G^{Se}G DNA dimers in order to compare the absorption peaks and extinction coefficient between ^{Se}G and G. The structure of the ^{Se}G^{Se}G dimer is presented in Figure 7.3 (3). Similarly to ^{Se}TTP, the UV absorption spectra of the ^{Se}G-containing DNA dimers have two peaks at ~260 nm and 360 nm, whereas the native GG DNA dimer does not absorb at 360 nm, with only a peak at 254 nm. The UV spectra of GG dimer, ^{Se}GG dimer and ^{Se}G^{Se}G dimer are illustrated in Figure 7.5. The colour of the Se-containing dimers is also yellow similar to ^{Se}TTP. The extinction coefficient of the 260 nm peak of the GG dimer (1.22×10^4 M^{-1} cm^{-1})[4] is roughly twice that of the 260 nm peak of the ^{Se}G-containing DNA dimers (5.3×10^3 M^{-1} cm^{-1})[2], whereas it is about half that of the 360 nm peak (2.3×10^4 M^{-1} cm^{-1})[2]. Again, the origin of the 360 nm peak of the Se-containing DNA dimers has not yet been determined.

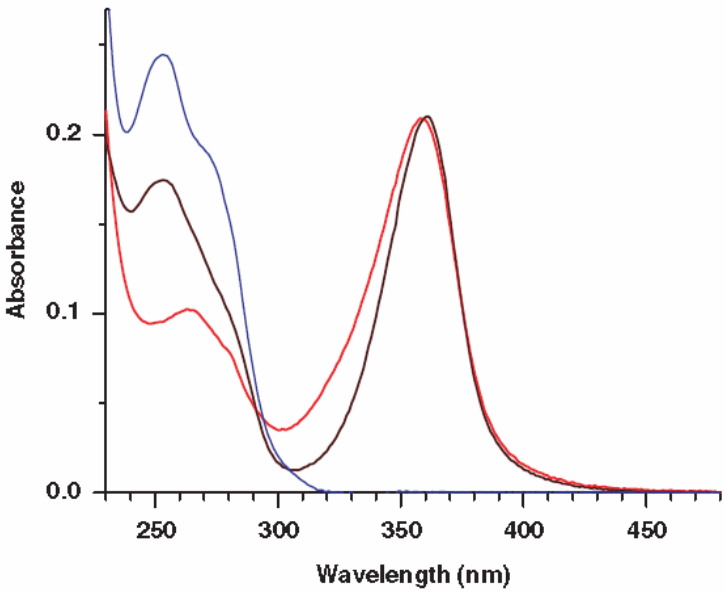


Figure 7.5 The UV absorption spectra of GG dimer (blue line), ^{Se}GG dimer (black line) and ^{Se}G^{Se}G dimer (red line). Reproduced from reference 2.

^{Se}G-containing DNA was also synthesised and studied by Salon *et al.*[2]. The results showed that ^{Se}G-containing DNA is relatively stable in aqueous solution at 60 °C over one hour. Furthermore, the crystal structure of ^{Se}G-containing DNA was seen to be similar to that of the native DNA duplex. Finally, the ^{Se}G and cytosine (C) form a base pair similar to the native G-C base pair with the Se atom also forming a hydrogen bond with C.

A number of levels of theory have been used to investigate the electronic structures and photophysical properties of the DNA canonical bases T and G, the structures of which are shown in Figure 7.2. Ground-state calculations have been performed using MP2[5-10], HF[8,11] and DFT[7,12-15], whilst CIS[7,8,11,13], TDDFT[5,6,12,14,16-18], CASPT2 and CASSCF[9,10,15,19-23] have been used to study photophysics and excited-state properties.

7.1.1 Equilibrium Geometries

A summary of the optimised ground - and first excited-state equilibrium geometries of T and G calculated at different levels of theory is shown in Table 7.1.

Geometry	Method	Reference
Optimised ground-state equilibrium geometry of T		
planar	MP2/6-311G(d,p)	5
planar except the methyl group hydrogens	HF/4-31G	11
Optimised first excited-state equilibrium geometry of T		
non-planar (boat like)	PCM/(TD)PBE0/6-31G(d)	12
non-planar	CIS/4-31G	11
non-planar	CASSCF	9
Optimised ground-state equilibrium geometry of G		
planar except the amino group	MP2/6-311G(d,p)	5
planar except the two hydrogens in amino group	B3LYP/TZVP	14
planar except the amino group	B3LYP/cc- pVDZ	13
planar except the amino group	PCM/B3LYP/cc-pVDZ	13
Optimised first excited-state equilibrium geometry of G		
non-planar (high distortion on the six-membered ring)[a]	(TD)B3LYP/TZVP	14
non-planar (high distortion on the six-membered ring)	CIS/cc-pVDZ	13
non-planar (high distortion on the six-membered ring)[b]	PCM/CIS/cc-pVDZ	13
non-planar (high distortion on the six-membered ring)[a]	CIS/4-31G mixed basis set[c]	8
non-planar (high distortion on the six-membered ring)[a]	CIS/6-311G(d,p)	24
non-planar (high distortion on the six-membered ring)[b]	CASSCF/6-31G(d)	10

Table 7.1 A summary of the optimised ground - and first excited-state equilibrium geometries of T and G calculated at various levels of theory. a means the optimised first excited-state equilibrium geometry of G calculated at these levels of theory has the same distortion direction, which is opposite to the distortion direction at those levels of theory labelled by b. c means CIS/4-31G basis set for all the atoms except the nitrogen atom of the amino group with 6-311+G(d) basis set.

The optimised ground-state equilibrium geometries of T and G in the gas phase were calculated using MP2/6-311G(d,p) by Shukla *et al.*[5]. Their results showed that T and G in ground-state were planar except the amino group in G. However, they did not mention to what extent the amino group in G was out of the plane in their paper. Except for the amino group in G, the calculated ground-state geometries of T and G were consistent with the crystal structures of T[25] and G[26], implying the MP2/6-311G(d,p) can describe the ground-state geometries of T and G quite satisfactorily.

Gustavsson *et al.*[12] showed the optimised first excited-state equilibrium geometry of T in solution, calculated using polarisable continuum model (PCM) and TDDFT with

PBE0/6-31G(d) (from now on called PCM/(TD)PBE0/6-31G(d)), to be non-planar, with the pyrimidine ring instead having a boat-like structure, which is similar to that of uracil (U) shown in Figure 7.6.

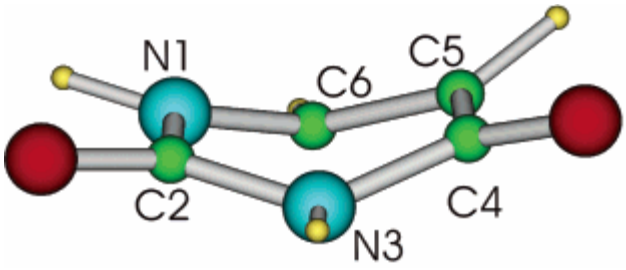
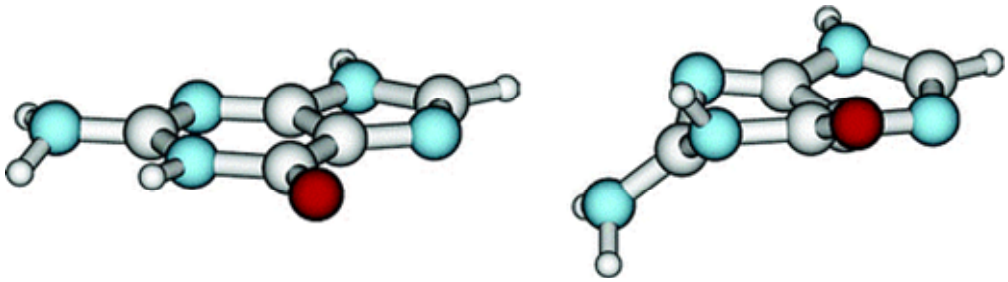


Figure 7.6 The optimised first excited-state equilibrium geometry of uracil calculated using PCM/(TD)PBE0/6-31G(d). Reproduced from reference 12.

Marian[14] calculated the optimised ground-state equilibrium geometry of G in the gas phase using B3LYP/TZVP and the optimised first excited-state equilibrium geometry using (TD)B3LYP/TZVP. Their result showed that ground-state structure of G is planar, except for the two amino hydrogen atoms which lie out of the ring plane, as shown in Figure 7.7 (1), which is similar to that obtained from MP2/6-311G(d,p) by Shukla *et al.*[5] and is consistent with the crystal structure of G[26], except for amino hydrogens. The first excited-state structure of G was found to be non-planar and highly distorted, as shown in Figure 7.7 (2).



(1) (2)

Figure 7.7 The optimised (1) ground-state structure calculated using B3LYP/TZVP and (2) first excited-state geometry calculated using (TD)B3LYP/TZVP of G in the gas phase. Reproduced from reference 14.

Mennucci *et al.*[13] calculated the optimised ground-state equilibrium geometry of G in the gas phase using B3LYP/cc-pVDZ and in solution using PCM/B3LYP/cc-pVDZ. Their results showed that the amino group is non-planar in both gas phase and

solution, which is consistent with that obtained from MP2/6-311G(d,p)[5], B3LYP/TZVP[14] in the gas phase and different to the crystal structure of G in which the amino group is in the same plane as the six - and five-membered rings[26]. The first excited-state geometry of G was also calculated using CIS/cc-pVDZ in the gas phase and PCM/CIS/cc-pVDZ in soluttion[13]. It was found to be non-planar with geometrical distortion on the six-membered ring in both gas phase and solution. The distortion in the solvated molecule is in the opposite direction to that in the gas phase obtained from (TD)B3LYP/TZVP[14], as shown in Figure 7.8.

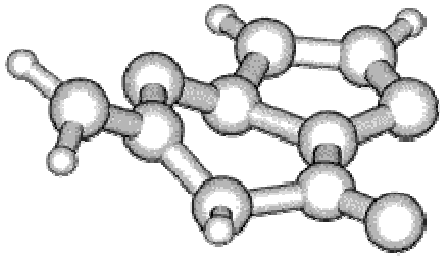


Figure 7.8 The optimised first excited-state geometry of G in solution calculated using PCM/CIS/cc-pVDZ. Reproduced from reference 13.

Shukla *et al.*[11] performed the ground-state equilibrium geometry optimisation, using HF/4-31G, and the first excited-state geometry optimisation using CIS/4-31G, of T in the gas phase. They found the ring of ground-state structure in the gas phase is planar (except the methyl group hydrogens), which is consistent with the crystal structure[25], whereas the first excited-state geometry of T is non-planar.

Shukla *et al.*[8] calculated the optimised first excited-state equilibrium geometry of G in the gas phase, using CIS/4-31G basis set for all the atoms, except the nitrogen atom of the amino group with 6-311+G(d) basis set; from now on this will be called the 4-31G mixed basis set. Their results showed that first excited-state geometry of G in the gas phase is non-planar and distorted, particularly in the six-membered ring region, as shown in Figure 7.9, which is consistent with that calculated using (TD)B3LYP/TZVP[14].

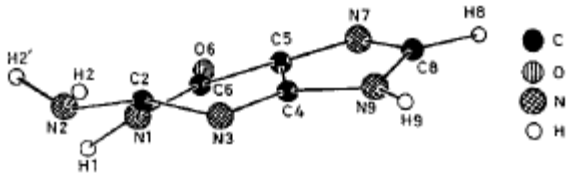


Figure 7.9 The optimised first-excited geometry of G in the gas phase calculated using CIS/4-31G mixed basis set. Reproduced from reference 8.

Zhang *et al*.[24] calculated the optimised first excited-state equilibrium structure of G in the gas phase using CIS/6-311G(d,p) and found it to be highly non-planar. The most notable feature is the dramatic out-of-plane distortion of the C_2 atom and the pyramidalization at the N_1 and C_5 atoms in the pyrimidine ring, shown in Figure 7.10. This is consistent with using (TD)B3LYP/TZVP[14] and CIS/4-31G mixed basis set[8].

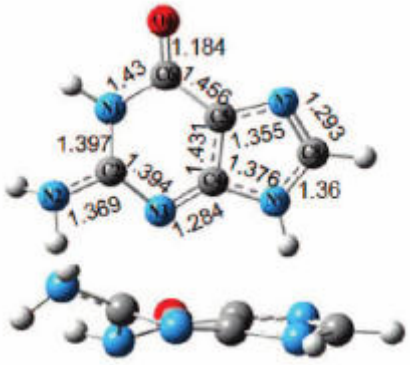


Figure 7.10 The atom numbering, and the top and side view of the optimised first excited-state geometry of G in the gas phase calculated using CIS/6-311G(d,p) basis set. Reproduced from reference 24.

Perun *et al*.[9] calculated the optimised first-excited equilibrium geometry of T in the gas phase at the CASSCF level of theory and found it to have a slightly out of plane distorted structure. The distortion mainly involves the $C_5C_4N_3$ (for atom labelling see Figure 7.2) moiety.

Chen *et al*.[10] calculated the optimised first-excited equilibrium geometry of G in the gas phase using CASSCF/6-31G(d). The result showed a distorted structure with the distortion mainly of the bending of the amino group and the atom connected to the

amino group being pyramidalized, shown in Figure 7.11, which is consistent with that of G in solution obtained by PCM/B3LYP/cc-pVD[13].

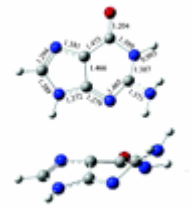


Figure 7.11 The top and side view of the optimised first excited-state equilibrium geometry of G calculated using CASSCF/6-31G(d,p). Reproduced from reference 10.

From above, it can be seen that different levels of theory, including the highest level of theory, CASSCF, generate similar results for the optimised ground - and first excited-state geometries of T and G. In general, T and G are predicted to be planar in ground-state, but non-planar in the first excited-state.

7.1.2 Singlet Absorption Transitions

A summary of the calculated electronic singlet absorption transitions of T and G at different levels of theory is shown in Table 7.2 and 7.3, respectively.

Excited-state	Transition type	Wavelength/ nm	Oscillator strength	Method	Reference
1 2	n→π* π→π*	263 250	0 0.14	(TD)B3LYP/6-311++G(d, p)	5
1	-	273	-	(TD)B3LYP/6-31+(d,p)	16
1 2	n→π* π→π*	254 245	- 0.15	(TD)PBE0/6-31G(d)	12
1 2	n→π* π→π*	242 238	- 0.18	PCM/(TD)PBE0/6-31G(d)	12
1 2	π→π* n→π*	245 239	0.21 -	PCM/(TD)PBE0/6-311+G(d,p)	12
1 2	π→π* n→π*	249 240	0.20 -	PCM/(TD)PBE0/6-311+G(2d,2p)	12
1 2	n→π* π→π*	- 261 (scaled)	0 0.50	CIS/4-31G	11
1 2	n→π* π→π*	282 254	0 0.17	CASPT2	22
N/A N/A	- -	264 206	- -	Experiment	27

Table 7.2 A summary of the electronic singlet absorption transitions of T calculated at different levels of theory and obtained from experiment.

Excited-state	Transition type	Wavelength/ nm	Oscillator strength	Method	Reference
1 3	π→π* π→π*	254 239	0.12 0.22	(TD)B3LYP/6-311++G(d,p)	5
1	-	282	-	(TD)B3LYP/6-31++G(d,p)	16
1	π→π*	274	0.22	(TD)B3LYP/TZVP	14
1	π→π*	239	0.15	(TD)B3LYP/6-311G(d,p)	24
1 2 3	π→π* n→π* π→π*	279 (scaled) 274 (scaled) 246 (scaled)	0.34 0.002 0.10	CIS/4-31G mixed basis set	8
1	π→π*	270 (scaled)	0.28	CIS/6-311G(d,p)	24
1 2 3	π→π* π→π* n→π*	275 (scaled) 248 (scaled) 237 (scaled)	0.24 0.58 0.02	SCRF-PCM/CIS/4-31G mixed basis set	8
1	π→π*	259	0.20	CASPT2/6-31G	10
N/A N/A	- -	275 246	- -	Experiment	28

Table 7.3 A summary of the electronic singlet absorption transitions of G calculated at different levels of theory and obtained from experiment.

Shukla *et al.*[5] performed calculations on the electronic singlet absorption transitions of T and G in the gas phase using (TD)B3LYP/6-311++G(d,p). Their results showed that the first transition of T is an n to π* transition and occurs at 263 nm (oscillator strength f = 0), whereas the second transition is a π to π* transition and occurs at 250 nm (f = 0.14). The calculated 250 nm peak is blue-shifted 14 nm compared with the experimental absorption spectrum of T in aqueous solution[27]. In contrast to T, the first transition of G is a π to π* transition and occurs at 254 nm (f = 0.12). Another calculated absorption peak of G with high oscillator strength occurs at 239 nm (f = 0.22), which is also assigned as a π to π* transition. These two peaks of G are blue-shifted compared with the experimental absorption spectrum of G in aqueous solution[28] with the largest deviation 21 nm. The oscillator strength of the 254 nm peak is roughly half of that of 239 nm peak, which is consistent with the experiment[28].

Varsano *et al.*[16] also studied the absorption spectra of T and G in the gas phase at (TD)B3LYP/6-31++G(d,p) level of theory. They only report the vertical energies of absorption transitions. They concluded that the first absorption transition occurs at 273 nm for T, which is 10 nm longer than the one from (TD)B3LYP/6-311++G(d,p)[5]. Similarly, the first absorption transition of G here is 282 nm, which is roughly 30 nm longer than the one from (TD)B3LYP/6-311++G(d,p)[5]. The differences suggest that the calculated absorption transitions are sensitive to the nature of the basis set.

The absorption spectrum of T in the gas phase was simulated at (TD)PBE0/6-31G(d) level of theory by Gustavsson *et al.*[12]. The calculated first absorption transition occurs at 254 nm and is an n to π* transition, the oscillator strength of which was not quoted in their paper. The transition type is consistent with that obtained from (TD)B3LYP/6-311++G(d,p)[5], whereas the transition peak is ~ 10 nm shorter than that from (TD)B3LYP/6-311++G(d,p)[5] and ~ 20 nm shorter than that from (TD)B3LYP/6-31++G(d,p)[16]. The second transition occurs at 245 nm (f = 0.15), which is consistent with that obtained from (TD)B3LYP/6-311++G(d,p)[5], and is a π to π* transition.

Gustavsson *et al.*[12] used the PCM/(TD)PBE0 combining with three different basis sets 6-31G(d), 6-311+G(d,p) and 6-311+G(2d, 2p) to predict the absorption spectrum of T in a solvated state. The first and second singlet absorption transitions, at 6-31G(d) level of theory, are quite similar to those in the gas phase (see above), with the absorption peaks being slightly shorter in solution. With increasing basis set, the transition types of the first and second transition are swapped compared with 6-31G(d): the first transition is π to π* in 6-311+G(d,p) and 6-311+G(2d, 2p), with the two corresponding absorption peaks being slightly longer. However, the absorption peaks of T in solution in all the three basis sets are shorter than those in the gas phase. They proposed that including solvation reduced the discrepancy between the calculated and experiment absorption peak (265 nm). However, as there is no calculated gas phase data at higher basis set in their paper, it is not clear whether the improvement in the position of the calculated absorption peak compared with experiment is from the inclusion of solvation or from the increased basis set used in solution calculations. The small (15 nm) blue-shift at 6-311+G(2d,2p) compared with experiment was attributed to the overstabilization of the π bonding orbital by (TD)PBE0 level of theory. It was suggested that it was important to include solvent effects with larger basis sets in order to generate the consistent result with experiment.

Marrian[14] also calculated the first singlet absorption transition of G in the gas phase at (TD)B3LYP/TZVP level of theory and found that it occurs at 274 nm (f = 0.22) and is a π to π* transition, which is 10 nm red-shifted compared with experiment[27]. The absorption peak is 20 nm longer than the one from (TD)B3LYP/6-311++G(d,p)[5] and ~10 nm shorter than the one from (TD)B3LYP/6-31++G(d,p)[16]. The oscillator strength here is almost twice that from TDDFT/B3LYP/6-311++G(d,p)[5], whereas the transition type is the same to the one from TDDFT/B3LYP/6-311++G(d,p)[5].

Zhang *et al.*[24] calculated the first singlet absorption transition of G in the gas phase at (TD)B3LYP/6-311G(d,p) level of theory and found that it occurs at 239 nm (f = 0.151), and is a π to π* transition. This peak is 15 nm shorter than the one from (TD)B3LYP/6-311++G(d,p)[5] and 36 nm blue-shifted compared with experiment[27].

From above, it can be seen that the prediction of the absorption spectrum at (TD)DFT level of theory is quite sensitive to the functional and basis set used. (TD)B3LYP/6-311++G(d,p), (TD)B3LYP/6-31++G(d,p), (TD)PBE0/6-31G(d) and (TD)B3LYP/TZVP generate moderately satisfactory absorption spectra of T and G with a small shift less than 40 nm compared with the corresponding experimental absorption spectra of T and G in aqueous solution.

The absorption transitions of T in the gas phase were also calculated using CIS/4-31G[11]. The result showed the first transition is an n to π* transition with oscillator strength 0, which is consistent with that obtained from (TD)B3LYP/6-311++G(d,p)[5] and (TD)PBE0/6-31G(d)[12]. The second transition of T occurs at 261nm (scaled, f = 0.5), which is only 3 nm shorter than that from experiment[27], indicating the CIS/4-31G level of theory can predict very satisfactory absorption peak of T. The oscillator strength of this transition is 3 times bigger than that obtained from (TD)B3LYP/6-311++G(d,p)[5] and (TD)PBE0/6-31G(d)[12]. The second transition was predicted to be a π to π* transition, which is consistent with that obtained from (TD)B3LYP/6-311++G(d,p)[5] and (TD)PBE0/6-31G(d)[12].

Shukla *et al*.[8] calculated the singlet absorption transitions of G in the gas phase using CIS/4-31G mixed basis set and in solution using the PCM model of self-consistent reaction field (SCRF) theory, with the same method and basis set as the gas phase. The first transition in the gas phase occurs at 279 nm (f = 0.345) and is a π to π* transition, of which the transition type is the same to that obtained from (TD)B3LYP/TZVP[14] and (TD)B3LYP/6-311++G(d,p)[5]. The first transition peak is 4 nm longer than experiment[27]. The second transition in the gas phase occurs at 274 nm (scaled, f ≈ 0) and is an n to π* transition, whereas the third transition occurs at 246 nm (scaled, f = 0.104) and is a π to π* transition. The calculated 246 nm peak corresponds exactly to experiment[28]. The first transition in solution occurs at 275 nm (scaled, f = 0.244), which is exactly the same to the experimental value[28]. This transition is also a π to π* type. Upon solvation, the transition types and oscillator strengths of the second and third transitions interchange, as a consequence, the second transition of solvated G occurs at 248 nm (scaled, f = 0.585) and is a π to π*

transition, the peak of which is 2 nm longer than the experiment[28]. The calculated third transition occurs at 237 nm (scaled, f = 0.020) and is an n to π* transition. All the calculated transition peaks in solution are a little blue-shifted compared with those in the gas phase. It can be seen that the SCRF-PCM/CIS/4-31G predicts the absorption spectrum of G satisfactorily compared with the experiment[28] and it is important to include the effect of solvation in order to predict more accurate absorption spectrum.

Zhang *et al*.[24] calculated the first singlet absorption transition of G in the gas phase using CIS/6-311G(d,p) and found that it occurs at 270 nm, and is a π to π* transition. This peak is 9 nm shorter than the one obtained from CIS/4-31G mixed basis set[8] and 5 nm blued-shifted compared with the experiment[27].

From the above, it can be concluded that CIS/4-31G and CIS/6-311G(d,p) levels of theory predict the absorption spectra of T and G successfully, with a very small shift less than 10 nm compared with the corresponding experimental absorption spectra. CIS predicts more accurate absorption spectra compared with the experiment than (TD)DFT. It is important to including solvation in both the (TD)DFT and CIS methods in order to predict accurate absorption spectra.

Lorentzon *et al.*[22] calculated the first two singlet absorption transitions of T in the gas phase at CASSCF and CASPT2 levels of theory. Their results obtained from CASPT2 showed the first transition occurs at 282 nm (f ≈ 0), and is an n to π* transition. The transition type and oscillator strength are the same to those obtained from (TD)B3LYP/6-311++G(d,p)[5], (TD)PBE0/6-31G(d)[12] and CIS/4-31G[11], whereas the peak wavelength is more than 20 nm longer than that from (TD)B3LYP/6-311++G(d,p)[5] and (TD)PBE0/6-31G(d)[12]. The second transition occurs at 254 nm (f = 0.17), and is a π to π* transition. Again, the transition type is the same to that obtained from (TD)B3LYP/6-311++G(d,p)[5], (TD)PBE0/6-31G(d)[12] and CIS/4-31G[11]. The accuracy of calculated absorption peak of T compared with the experiment[27] from low to high are: (TD)B3LYP/6-311++G(d,p)[5], (TD)PBE0/6-31G(d)[12] < CASPT2 < CIS/4-31G[11].

Chen *et al.*[10] calculated the first singlet absorption transition of G in the gas phase at CASSCF and CASPT2 levels of theory with 6-31G(d) basis set. Their result obtained from CASPT2/6-31G(d) showed that the transition occurs at 259 nm (f = 0.2), and is a π to π* transition. This 259 nm peak is similar to that calculated using TDDFT/B3LYP/6-311++G(d,p)[5] and 16 nm shorter than that from the experiment[28].

Therefore, the accuracy of the absorption spectra of T and G predicted at CASSCF and CASPT2 levels is similar to that using TDDFT/B3LYP/6-311++G(d,p). Taking the results from the different levels of theory together, the first absorption transition of T in the gas phase is predicted to be a forbidden n to π* transition, and the second transition is predicted to be a π to π* transition. The first absorption transition of G in the gas phase and in solution is predicted to be a π to π* transition.

7.1.3 Singlet Emission Transitions

A summary of the calculated electronic singlet emission transitions of T and G at different levels of theory is shown in Table 7.4.

Excited-state	Transition type	Wavelength/ nm	Oscillator strength	Method	Reference
T					
1	π→π*	342	> 0.20	PCM/(TD)PBE0/6-31G(d)	12
1	π→π*	353	> 0.20	PCM/(TD)PBE0/6-311+G(d,p)	12
1	π→π*	358	> 0.20	PCM/(TD)PBE0/6-311+G(2d,2p)	12
1	-	312 (scaled)	0	CIS/4-31G	11
1	-	329	-	Experiment	12
G					
1	π→π*	309	-	(TD)B3LYP/TZVP	14
1	π→π*	452	0.07	(TD)B3LYP/6-311G(d,p)	24
1	-	475	0.15	CIS/4-31G mixed basis set	8
1	π→π*	390	0.22	CIS/6-311G(d,p)	24
1	-	449	0.15	SCRF-PCM/CIS/4-31G mixed basis set	8
1	π→π*	319	-	CASPT2/6-31G	10
N/A	-	335	-	Experiment	28

Table 7.4 A summary of the electronic singlet emission transitions of T and G calculated at different levels of theory and obtained from experiment.

The first singlet emission transition of T in solution, calculated using PCM/(TD)PBE0[12], is 342 nm (6-31G(d)), 353 nm (6-311+G(d,p)) and 358 nm (6-311+G(2d,2p)) with high oscillator strength (> 0.2), and is assigned to a π to π* allowed transition. The calculated emission peak is red-shifted between 13 nm to 29 nm with increasing the basis set compared with the experimental fluorescence emission peak in aqueous solution (329 nm)[12]. The experimental fluorescence quantum yield of T is extremely low, 1.2 x 10^{-4} [29]. This is because that the lowest-energy conical intersection of T, arising from a crossing between the lowest singlet ππ* state and the ground-state and being accessible in a barrierless manner from the minimum of the lowest singlet ππ* state, provides a direct and fast pathway for the quenching of the population of the lowest singlet ππ* excited-state of T[9,23].

Marrian[14] also calculated the emission transition of G in the gas phase at (TD)B3LYP/TZVP level of theory. They predicted the first emission band to occur at 309 nm, which is 26 nm blue-shifted compared with the experimental fluorescence

spectrum of G in aqueous solution (with emission peak at 335 nm)[28], with a π to π* transition. The oscillator strength of this transition is not included in their paper. The experimental fluorescence quantum yield of G is also very low, 0.8 x 10^{-4} [29], due to the fast and efficient radiationless decay of G through a barrierless excited-state pathway leading to an out of plane deformation type conical intersection with the ground-state[15,19,20].

Zhang *et al*.[24] predicted the first singlet emission transition of G in the gas phase, using (TD)B3LYP/6-311G(d,p), to occur at 452 nm (f = 0.07), and to be a π to π* transition. This predicted wavelength is red-shifted more than 100 nm compared with the experiment[28] and nearly 150 nm longer than that obtained from (TD)B3LYP/TZVP[14]. Zhang *et al*. suggested that the 452 nm peak corresponds to the experimental fluorescence shoulder around 450 nm in the oxygen-rich aqueous G solution measured by Santhosh *et al*.[30]. However, Santhosh *et al*.[30] attributed the origin of this 450 nm shoulder to the first triplet excited-state of G, which became allowed due to the interaction between the G molecule and oxygen in aqueous solution. The interpretation of Zhang *et al.* is questionable since their calculation was performed in the gas phase and did not involve interaction with oxygen.

The first singlet emission transition of T in the gas phase was also calculated using CIS/4-31G[11] and found to occur at 312 nm (scaled) with oscillator strength 0. This transition is 17 nm blue-shifted compared with the experiment[12], in contrast to that obtained from PCM/(TD)PBE0/6-31G(d) (red-shifted 13 nm)[12]. From above, it can be seen that the accuracy between the predicted emission transition peak of T and experiment is similar between using CIS/4-31G and PCM/(TD)PBE0/6-31G(d).

Shukla *et al*.[8] calculated the first singlet emission transition of the two tautomers of G: keto-9NH and keto-7NH, in the gas phase using CIS/4-31G mixed basis set and in solution using SCRF-PCM model with the same method and basis set to the gas phase. From now on, if it is not specified, all the results are for keto-9NH of G. The transition occurs at 475 nm (scaled, f = 0.15) in the gas phase and 449 nm (scaled, f = 0.152) in solution. The peak of this transition is much longer than that obtained

from experiment (335 nm)[28]. Shukla *et al*. suggested this calculated 449 nm peak corresponds to the experimental fluorescence shoulder around 450 nm[30], and the 335 nm experimental peak[28] originates from the keto-N7H tautomer. Their results showed the first transition of the keto-N7H tautomer in solution occurs at 333 nm (scaled, f = 0.292), which supports his suggestion.

Zhang *et al*.[24] predicted the first singlet emission transition of G in the gas phase, using CIS/6-311G(d,p), to occur at 390 nm (f = 0.220), and to be a π to π* transition. The calculated peak is significantly shorter than the calculated 449 nm peak using CIS/4-31G mixed basis set[8] and red-shifted 55 nm compared with the experiment[28].

From the above, it can be seen that the (TD)DFT and CIS levels of theory for predicting the fluorescence spectra are quite sensitive to functional and basis set. In addition, the predicted oscillator strengths are also sensitive to functional and basis set.

Chen *et al*.[10] calculated the first singlet emission transition of G in the gas phase at CASSCF and CASPT2 levels of theory with 6-31G(d) basis set. Their result obtained from CASPT2/6-31G(d) showed that the transition occurs at 319 nm, 16 nm shorter than the experiment,[28] and is a π to π* transition.

From above, it can be concluded that (TD)B3LYP and CIS with different basis sets generate the first excited-state equilibrium geometriess of T and G similarly to CASSCF and CASPT2. In addition, CASSCF and CASPT2 with 6-31G(d) basis set generate moderately accurate results of absorption and emission transitions of T and G compared with experiment, and are slightly better than (TD)B3LYP/6-311++G(d,p)[5] and (TD)PBE0/6-31G(d)[12]. However, CASSCF and CASPT2 levels of theory are computationally expensive. Therefore, CASSCF and CASPT2 were not used in this work. The first excited-state, absorption and emission transitions of the natural bases and Se-base analogues were calculated using (TD)B3LYP and CIS methods in this work.

Although there have been numerous theoretical studies performed on the natural bases T and G, there has been limited effort devoted to the theoretical investigation of the structural and spectral properties of ^{Se}T and ^{Se}G.

In order investigate the potential of ^{Se}T as a component of nanowires, Vazquez-Mayagoitia *et al.*[31] calculated the optimised ground-state equilibrium structures of T and ^{Se}T in the gas phase, without imposing geometrical restrictions, at MP2/6-31G(d,p) level. Their results showed that the optimised bond lengths and bond angles of T and ^{Se}T are very similar except for the adjacent environment of the C_4-O/Se bond (for atom numbering see Figure 7.2). C_4-Se is 1.782 Å, which is 0.55 Å longer than C_4-O. They suggested the result was expected on the basis of the larger size of van der Waals radius of Se (2.0 Å) relative to O (1.4 Å)[32]. The bond lengths N_3-C_4 and C_4-C_5 connecting to C_4- Se in ^{Se}T decrease ~0.02 Å compared with the corresponding ones in T. The rest of the bond lengths are mainly unchanged. Se alters the bond angles only slightly with the differences of bond angles between T and ^{Se}T less than 0.8 °. Therefore, they drew the conclusion that replacing the O atom at 4-position of T with Se did not affect the structural properties.

^{Se}G is reported as an effective inhibitor to grow several mouse tumors[33], in order to compare the structure difference with G, Leszczynski[34] calculated the optimised ground-state equilibrium molecular geometries of G and ^{Se}G at HF level with valence double-zeta basis sets augmented by d and p polarization functions (DZP) with C_1 symmetry. His results showed similarities between the molecular structures of G and ^{Se}G. The calculated bond lengths and bond angles of the G and ^{Se}G are very similar except for those near the C_6-O/Se (for atom numbering see Figure 7.2). C_6-O is 1.194 Å in G, whereas C_6-Se is1.819 Å in ^{Se}G that is ~1.5 times as big as that in G. C_6-N_1 and C_6-C_5 are the two bond lengths most sensitive to Se, which decrease 0.03 Å in ^{Se}G compared with the corresponding ones in G. The differences of bond lengths and bond angles of the five-membered ring are not significant between G and ^{Se}G. The dipole moment of ^{Se}G is 7.96 D, which is a little bigger than that of G (6.14 D).

There are no previous reports of the calculated optimised first excited-state equilibrium geometries, the absorption and emission transitions of ^{Se}T and ^{Se}G.

In this chapter, the experimentally measured absorption spectra, steady-state fluorescence spectra and phosphorescence spectra of ^{Se}T*p and ^{Se}G*p (structures shown in Figure 7.1) are reported first and the potential of ^{Se}T*p and ^{Se}G*p as fluorescent probes is considered. The computational results performed at HF, MP2, B3LYP, (TD)B3LYP and CIS levels on the four bases T, ^{Se}T, G and ^{Se}G (structures shown in Figure 7.2) are then presented and compared with the experimental results.. The effects of replacement of an O atom by a Se atom in T and G on the molecular structures and spectral properties are elucidated.

7.2 Materials, Methods, and Computational Details

^{Se}T*p and ^{Se}G*p, shown in Figure 7.1, were used as received without further purification. They were dissolved in aqueous solution at room temperature. The concentration of ^{Se}T*p in aqueous solution was ~11 μM and that of ^{Se}G*p in aqueous solution was ~7.9 μM. For the measurement at 77 K, both were dissolved in 10 M LiCl solution.

UV-Vis absorption spectra were measured using a Varian Cary 300 UV-Vis spectrophotometer, in a matched pair of fused silica 1 cm path length cuvettes (Hellma).

Steady-state fluorescence was measured using a Fluoromax spectrofluorometer, as described in Chapter 3. The method of freezing ^{Se}T*p to 77 K was described in Section 3.4 of Chapter 3. The excitation spectra at room temperature and 77 K were both recorded with emission wavelengths from 270nm to 520nm, at 10 nm intervals. The emission spectra were measured with excitation wavelengths from 200 nm to 400 nm, at 10 nm intervals. All the excitation and emission spectra of ^{Se}T*p in LiCl at room temperature and 77 K presented here have been corrected for background

emission from the LiCl. The fluorescence quantum yield of ^{Se}G*p in aqueous solution at room temperature was calculated using comparing the fluorescence emission intensity excited at 340 nm with that of 2-aminopurine riboside (2APr), of which the quantum yield is known (0.68[35]), under exactly the same experimental conditions.

The computational methodology is discussed in Section 3.5 of Chapter 3. The keto forms of T, G, ^{Se}T and ^{Se}G have been used for the theoretical calculations, as they are the most stable forms in the electronic ground-state in both gas phase and in aqueous solution[31,36,37]. In addition, the keto tautomers of T and G are the major tautomer in the gas phase[8,38,39] and solution-phase[8,39,40]. For G, the keto-7NH tautomer is most stable in the gas phase, and keto-9NH tautomer is most stable in aqueous solution[8]. Therefore, keto amide tautomer of T, keto-N9H of G, and the corresponding Se-base analogues were chosen and investigated by computational calculations. As in previous studies,[7,11,41] the absorption and emission transition energies of all the four molecules obtained from CIS method are linearly scaled by a factor of 0.72, because CIS is equivalent to a calculation of HF quality in which the calculated transition energy is overestimated due to neglect of instantaneous electron correlation.

7.3 Results and Discussion

7.3.1 Experimental

7.3.1.1 UV-Vis Absorption Spectra of ^{Se}T*p and ^{Se}G*p at Room Temperature

The UV-Vis absorption spectra of ^{Se}T*p and ^{Se}G*p in aqueous solution at room temperature are shown in Figure 7.12(1), and those of T[27] and G[28] in aqueous solution from literatures are shown in Figure 7.12 (2) and (3). The absorption peaks of ^{Se}T*p and ^{Se}G*p in aqueous solution are summarised in Table 7.5, together with the literature values for T and G.

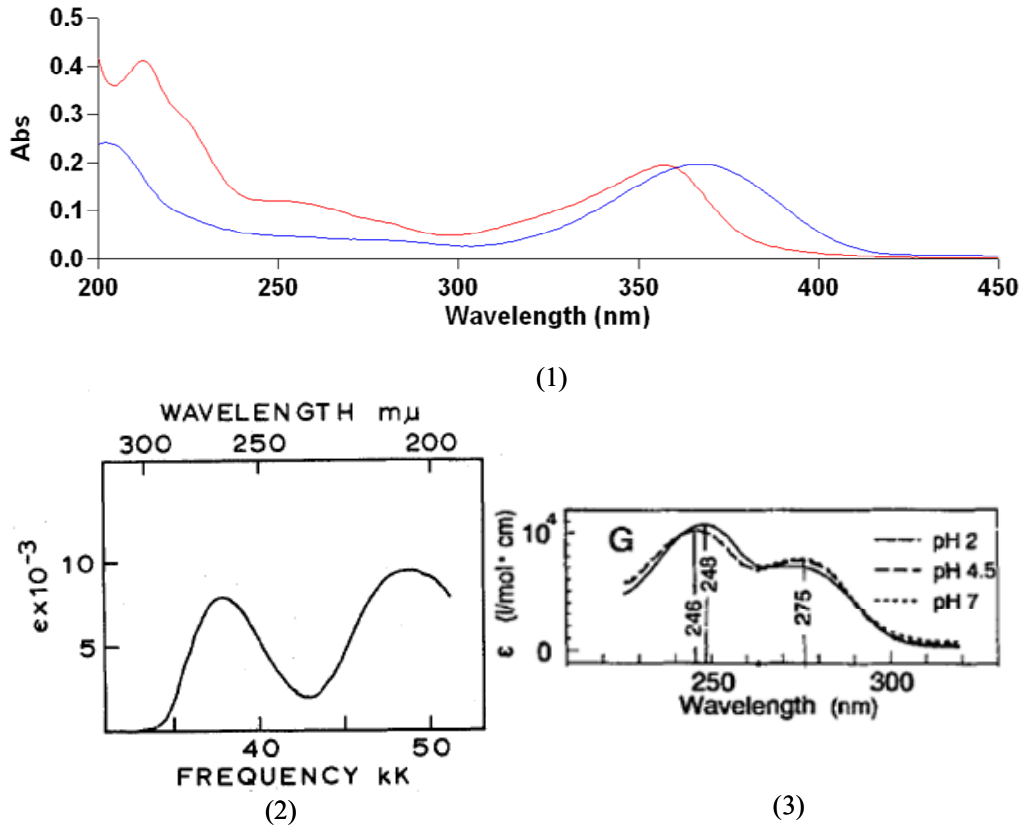


Figure 7.12 The UV-Vis absorption spectra of (1) ^{Se}T*p (blue line) and ^{Se}G*p (red line); (2) T, reproduced from reference 27 and (3) G, reproduced from reference 28; in aqueous solution at room temperature.

Sample	Absorption wavelength/ nm				Reference
^{Se}T*p	368	282 (S)	250 (S)	203	this work
T (pH = 7)	-	-	264	206	27
^{Se}G*p	357	279(S)	253 (S)	213	this work
G (pH = 7)	-	275	246	-	28

Table 7.5 Main absorption bands of ^{Se}T*p, T, ^{Se}G*p and G in aqueous solution at room temperature. S stands for shoulder on the absorption spectrum.

The main absorption peaks of ^{Se}T*p in aqueous solution at room temperature occur at 368 nm and 203 nm. The intensity of the peak at 368 nm (ε = 1.8 x 10^4 M^{-1} cm^{-1}) is slightly lower than that at 203 nm. In addition, the absorption spectrum of ^{Se}T*p has two shoulders at about 282 nm and 250 nm. The absorption peaks and extinction coefficients of ^{Se}T*p are consistent with those reported by Caton-Williams *et al.*[1] for ^{Se}TTP, as shown in Figure 7.4, indicating the extra two phosphates have almost no

influence on the absorption spectrum. The 368 nm absorption peak of ^{Se}T*p is shifted to longer wavelength by ~100 nm relative to the long wavelength absorption peak (264 nm) of T[27], whereas the 203 nm absorption peak of ^{Se}T*p is almost identical to that of T (206 nm)[27]. The 250 nm shoulder of ^{Se}T*p is quite similar to the long wavelength absorption peak of T[27].

The absorption spectrum of ^{Se}G*p in aqueous solution at room temperature has two main peaks, at 357 nm ($\varepsilon = 2.4 \times 10^4$ M^{-1} cm^{-1}) and 213 nm, as well as two shoulders at roughly 279 nm and 253 nm. The intensity of the 357 nm peak is much weaker than that of the 213 nm peak. Again the spectrum of ^{Se}G*p is consistent with that reported by Salon *et al*.[2] for the G^{Se}G dimer, as shown in Figure 7.5, implying that G in G^{Se}G dimer has little influence on the absorption spectrum of G^{Se}G. The 357 nm absorption peak of ^{Se}G*p is shifted to longer wavelength by ~80 nm relative to the long wavelength peak of G in aqueous solution[28]. The 253 nm shoulder of ^{Se}G*p is similar to the short wavelength peak of G in aqueous solution[28].

The long wavelength absorption peaks of both ^{Se}T*p and ^{Se}G*p are more than 80 nm longer compared with those of T and G, which is consistent with previous reports[1,2] that Se replacement of O in the natural bases shifts the UV absorption by about 100 nm to the red.

7.3.1.2 Fluorescence Spectra of ^{Se}T*p and ^{Se}G*p at Room Temperature

The fluorescence emission spectra of ^{Se}T*p and ^{Se}G*p in aqueous solution at room temperature recorded with an excitation wavelength of 370 nm (that is near the absorption peak of both ^{Se}T*p and ^{Se}G*p) are shown in Figure 7.13. The peak at about 420 nm in both fluorescence emission spectra is the water Raman band. The tiny peak at ~450 nm in spectrum of ^{Se}T*p comes from the tail of the Raman band, since it shifts with different excitation wavelengths. Therefore, ^{Se}T*p is not fluorescent, which is similar to T in aqueous solution which has an extremely low fluorescence quantum yield (1.2×10^{-4})[29].

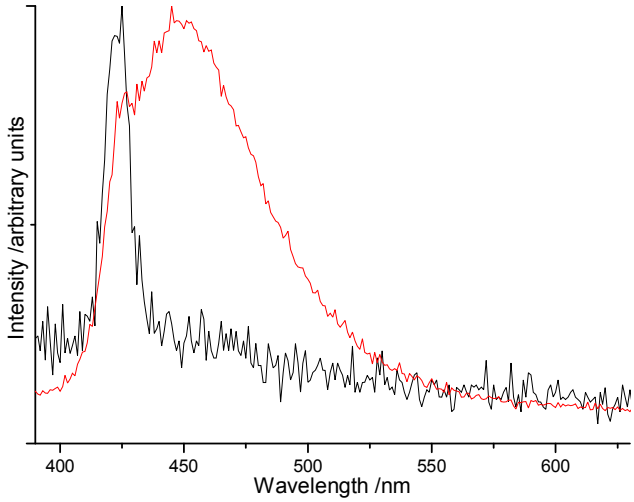


Figure 7.13 The fluorescence emission spectra of ^{Se}T*p (black line, intensity multiplied by a factor of ~13) and ^{Se}G*p (red line) in aqueous solution at room temperature with excitation wavelength 370 nm. The sharp peak at ~420 nm in the emission spectrum of ^{Se}T*p is due to the water Raman band.

^{Se}G*p is weakly fluorescent, with an emission at about 450 nm that is red-shifted ~120 nm compared with G in aqueous solution[28]. The approximate fluorescence quantum yield of ^{Se}G*p in aqueous solution at room temperature is 1.2 x 10^{-2} that is 150 times that of G in aqueous solution[29]. Although the fluorescence quantum yield of ^{Se}G*p is much bigger than that of G, it is still very small, which means that ^{Se}G*p is not very promising as a fluorescent base analogue.

7.3.1.3 Phosphorescence Spectrum of ^{Se}T*p at 77 K

The emission spectrum of ^{Se}T*p in LiCl was recorded at an excitation wavelength of 370 nm at room temperature and 77 K, shown in Figure 7.14. The emission spectrum of LiCl itself at this excitation wavelength at 77 K is also presented. The emission spectrum of ^{Se}T*p in LiCl at room temperature shows no detectable emission, whereas that at 77 K there appears to be a weak emission band with maximum intensity at about 585 nm. Therefore, the emission at 77 K is tentatively assigned as phosphorescence. The phosphorescence of T in ethylene glycol and water (1:1) at 77 K is observed at 430 nm[42]; therefore, Se replacement of O in T makes the peak of phosphorescence shift to the red by about 150 nm. The emission spectrum of ^{Se}T*p at

77 K has some other peaks in the region between 400 nm and 500 nm that may be due to scattered light from the frozen sample.

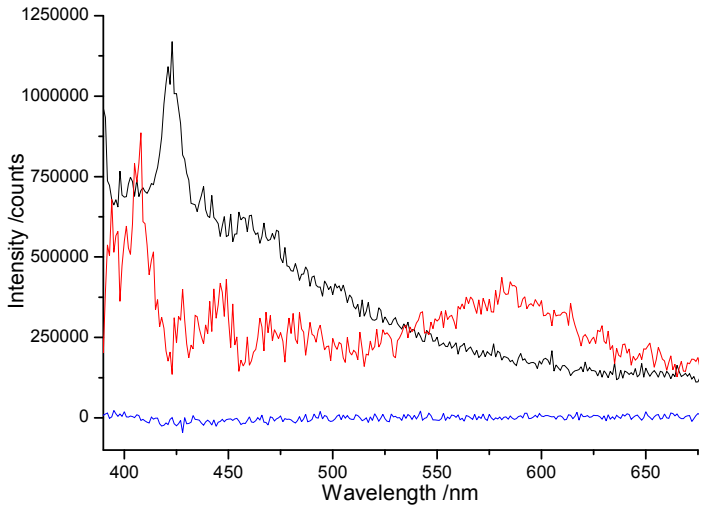


Figure 7.14 The emission spectra of ^{Se}T*p in LiCl with excitation wavelength 370 nm at room temperature (blue line), 77 K (red line) and that of LiCl itself at 77 K (black line).

The emission spectrum of ^{Se}G*p at 77 K could not be measured because of contamination or degradation of the sample during storage.

In conclusion, substitution of O by Se has a similar effect on the absorption and emission spectra of the natural bases T and G, causing the peaks of absorption, fluorescence and phosphorescence spectra to shift to longer wavelength by at least 80 nm. Although the ^{Se}T*p and ^{Se}G*p can be selectively excited and ^{Se}G*p can be selectively detected compared with the corresponding native bases, the low fluorescence emission quantum yield of ^{Se}G*p precludes its use as a fluorescent base analogue.

7.3.2 Calculations on T, ^{Se}T, G and ^{Se}G

7.3.2.1 Ground-state Equilibrium Geometry

The optimised ground-state equilibrium geometries of T, ^{Se}T, G and ^{Se}G in the gas phase were calculated at HF, B3LYP and MP2 levels of theory with 6-311+G(d,p)

basis set. Those in solution (water) were optimised using SCRF-PCM/MP2/6-311+G(d,p).

7.3.2.1.1 Calculated Bond Lengths between Different Levels of Theory

The calculated bond lengths of T and ^{Se}T in the gas phase using HF, B3LYP and MP2 with 6-311+G(d,p) basis set and in solution using SCRF-PCM/MP2/6-311+G(d,p) are presented in Table C1 in Appendix C, and those of G and ^{Se}G are presented in Table C2 in Appendix C. It can be seen that the calculated bond lengths of each molecule by HF are different to those from B3LYP and MP2 methods, with most of the bonds being shorter than the corresponding ones from B3LYP and MP2, due to the neglect of electron correlation by HF. The bond lengths of each molecule calculated using B3LYP are extremely similar to those from MP2, with the largest deviation of 0.02 Å. The calculated bond lengths of each molecule by PCM/MP2 in solution are similar to those from MP2 in the gas phase, except for C_4-O/Se, C_2-O and those with one side connecting with C_4 or C_2 in T and ^{Se}T; and C_6-O/Se, C_2-N and those with one side connecting with C_6 or C_2 in G and ^{Se}G.. This is probably because the polar bonds are affected by the polarity of the solvent.

7.3.2.1.2 Calculated Bond Angles between Different Levels of Theory

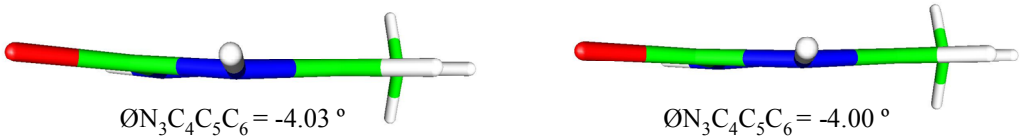
The calculated bond angles of T and ^{Se}T in the gas phase using HF, B3LYP, MP2 with 6-311+G(d,p) basis set and in solution using SCRF-PCM/MP2/6-311+G(d,p) are presented in Table C3 in Appendix C, and those of G and ^{Se}G are presented in Table C4 in Appendix C. The calculated bond angles of each molecule from HF are consistent with those from B3LYP and MP2 methods. The differences of calculated bond angles of each molecule between B3LYP, HF and MP2 are no more than 1.8 °. Upon solvation, the calculated bond angles from PCM/MP2 do not differ greatly from those in the gas phase, with a largest deviation of 1.6 °, except for the four bond angles centred on C_5, C_6 and N_1 atoms (C_5-C_6-N_1, C_5-C_6-Se, C_6-N_1-H and N_7-C_5-C_6) in ^{Se}G with largest deviation of 2.8 °, indicating solvation does not have significant effect on the calculated bond angle of molecules.

7.3.2.1.3 Planarity and Calculated Torsion Angles at Different Levels of Theory

The side views and torsion angles of T and ^{Se}T in the gas phase from B3LYP and HF calculations with 6-311+G(d,p) basis set are shown in Figures C1 and C2 in Appendix C. All the atoms of T and ^{Se}T, calculated using B3LYP and HF methods, are in the same plane (except for the two hydrogens in methyl group). The results for T obtained from HF/6-311+G(d,p) are consistent with those from calculated using HF/4-31G[11], and the structures calculated using HF/6-311+G(d,p) and B3LYP/6-311+G(d,p) are consistent with the crystal structure of T anhydrate[25].

The side views and torsion angles of G and ^{Se}G in the gas phase from B3LYP and HF calculations with 6-311+G(d,p) basis set are shown in Figures C3 and C4 in Appendix C. As for T and ^{Se}T, all the atoms of G and ^{Se}G calculated using B3LYP and HF methods are coplanar except for the two hydrogens in amino group. The results for G obtained from B3LYP/6-311+G(d,p) are consistent with those optimised using B3LYP/6-31G(d,p)[7], B3LYP/6-311G(d)[15], B3LYP/cc-pVDZ[13] and B3LYP/TZVP[14]. The calculated structures are consistent with the crystal structure of G monohydrate[26].

The side view and torsion angles of T and ^{Se}T in the gas phase at MP2/6-311+G(d,p) level of theory are shown in Figure 7.15. It can be seen that the six-membered rings are not flat, but a little buckled with the maximum torsion angle ~11 ° for T and ~9 ° for ^{Se}T. This calculated structure of T in the gas phase is different to the crystal structure of T anhydrate[25], which is nearly planar, and also differs from the planar structures predicted by MP2/cc-PVDZ[9] and MP2/6-311G(d,p)[5]. The structure of T obtained using MP2/6-311+G(d,p) in this work does not have C_S symmetry constraint in running the calculation, whereas that using MP2/cc-PVDZ level[9] is under the C_S symmetry constraint. However, MP2/6-311G(d,p)[5] also does not have C_S symmetry constraint, suggesting that the diffuse function accounts for the buckling of T.



ØN$_3$C$_4$C$_5$C$_6$ = -4.03 °
ØC$_4$C$_5$C$_6$N$_1$ = 3.17 °
ØC$_5$C$_6$N$_1$C$_2$ = -6.12 °
ØC$_6$N$_1$C$_2$N$_3$ = 9.09 °
ØN$_1$C$_2$N$_3$C$_4$ = -10.96 °
ØC$_2$N$_3$C$_4$C$_5$ = 8.70 °
(1)

ØN$_3$C$_4$C$_5$C$_6$ = -4.00 °
ØC$_4$C$_5$C$_6$N$_1$ = 2.90 °
ØC$_5$C$_6$N$_1$C$_2$ = -4.94 °
ØC$_6$N$_1$C$_2$N$_3$ = 7.38 °
ØN$_1$C$_2$N$_3$C$_4$ = -9.42 °
ØC$_2$N$_3$C$_4$C$_5$ = 7.96 °
(2)

Figure 7.15 The side view and torsion angles of optimised ground-state equilibrium geometries of (1) T and (2) ^{Se}T in the gas phase calculated at MP2/6-311+G(d,p) level of theory. Green stands for carbon, blue for nitrogen, white for hydrogen, and red for oxygen.

The side view and torsion angles of G and ^{Se}G in the gas phase at MP2/6-311+G(d,p) level of theory are shown in Figure 7.16. The six-membered ring of G and ^{Se}G is slightly buckled with the maximum torsion angle ~4 ° for both G and ^{Se}G, whereas the five-membered ring of G and ^{Se}G is almost planar. The structure of G is different to that obtained from MP2/6-311G(d,p)[5] and the crystal structure of G monohydrate[26], that is essentially planar. This again suggests that diffuse function is the reason for the buckling.

Six-member ring
ØC$_5$C$_6$N$_1$C$_2$ = -3.87 °
ØC$_6$N$_1$C$_2$N$_3$ = 2.89 °
ØN$_1$C$_2$N$_3$C$_4$ = -1.82 °
ØC$_2$N$_3$C$_4$C$_5$ = 2.82 °
ØN$_3$C$_4$C$_5$C$_6$ = -4.42 °
ØC$_4$C$_5$C$_6$N$_1$ = 4.23 °
Five-member ring
ØC$_8$N$_7$C$_5$C$_4$ = -0.97 °
ØN$_7$C$_5$C$_4$N$_9$ = 1.20 °
ØC$_5$C$_4$N$_9$C$_8$ = -0.92 °
ØC$_4$N$_9$C$_8$N$_7$ = 0.38 °
ØN$_9$C$_8$N$_7$C$_5$ = 0.35 °
(1)

Six-member ring
ØC$_5$C$_6$N$_1$C$_2$ = -1.20 °
ØC$_6$N$_1$C$_2$N$_3$ = 0.40°
ØN$_1$C$_2$N$_3$C$_4$ = -0.88 °
ØC$_2$N$_3$C$_4$C$_5$ = 2.63 °
ØN$_3$C$_4$C$_5$C$_6$ = -3.69 °
ØC$_4$C$_5$C$_6$N$_1$ = 2.53°
Five-member ring
ØC$_8$N$_7$C$_5$C$_4$ = -0.63 °
ØN$_7$C$_5$C$_4$N$_9$ = 0.84 °
ØC$_5$C$_4$N$_9$C$_8$ = -0.68 °
ØC$_4$N$_9$C$_8$N$_7$ = 0.33 °
ØN$_9$C$_8$N$_7$C$_5$ = 0.18 °
(2)

Figure 7.16 The side view and torsion angles of the optimised ground-state equilibrium geometries of (1) G and (2) ^{Se}G in the gas phase calculated at MP2/6-311+G(d,p) level. Green stands for carbon, blue for nitrogen, and white for hydrogen.

All the atoms of T and ^{Se}T in solution calculated by SCRF-PCM/MP2/6-311+G(d,p) are in the same plane (except for the two hydrogens in methyl group), as shown in

Figure C5 in Appendix C. This is consistent with the crystal structures of T monohydrate[43]. This is also in agreement with the gas phase structures from HF and B3LYP, which are planar.

The side view and torsion angles of G and ^{Se}G in solution calculated at SCRF-PCM/MP2/6-311+G(d,p) level are shown in Figure C6 in Appendix C. The six-membered ring of G and ^{Se}G is almost planar, with torsion angles that are slightly smaller than the corresponding ones in the gas phase, which is similar to the planar crystal structure of G monohydrate[26].

7.3.2.1.4 Intramolecular Basis Set Deficiency of MP2/6-311+G(d,p)

When using MP2/6-311+G(d,p) to calculate the optimised ground-state equilibrium geometries of T, ^{Se}T and ^{Se}G in the gas phase, the optimisations first converged to planar structures with one imaginary frequency for T and ^{Se}T, and two imaginary frequencies for ^{Se}G. Therefore, these structures are not minima on the potential energy surface (PES), but are saddle points. This observation is consistent with the results of Asturiol *et al*.[44]. They found when adding diffuse function to MP2/6-31G(d) and MP2/6-311G(d), the optimised planar structures of T and U have one imaginary frequency, whereas at the CASSCF level, no imaginary frequencies have been found in any case. They suggested that this problem could be associated with the electronic excitations to high energy virtual orbitals with large diffuse character. When no diffuse functions are included in the basis set, the planar structures correspond to true minima[44]. Asturiol *et al.* have demonstrated the origin of this pitfalls to be rooted in intramolecular basis set deficiencies, which eventually lead to intramolecular basis set superposition error, BSSE[44]. The BSSE arises due to the incompleteness of the basis set centred on the atomic positions and to which MP2 is known to be susceptible[44]. A similar buckling in aromatic systems has been observed for MP2 calculations on systems such as benzene, non-rigid cyclic systems, DNA and RNA nucleobases, polycyclic aromatic, and dihydrogen bonded systems[44-49]. The most widely used method to correct the BSSE effect is the counterpoise (CP)

method[48,50-52]. Asturiol *et al*.[44] showed that CP method supplied a proper assessment and correction to the BSSE problem caused by MP2 method.

In this work, in order to find the true minimum on the PES, using a saddle point as a starting point, the Cartesian co-ordinates of the first imaginary frequency were added to the Cartesian co-ordinates of the saddle point, and then the new co-ordinates were used to rerun the geometry optimisation and frequency calculations until no imaginary frequency occurred. The optimised ground-state equilibrium geometries of the four molecules in the gas phase calculated using MP2/6-311+G(d,p) (with no imaginary frequencies) are slightly buckled (as shown above in Figure 7.15 and 7.16). The torsion angles of the six-membered ring of G and ^{Se}G in the gas phase are smaller than those of T and ^{Se}T, indicating the intramolecular basis set deficiency has a smaller effect on G and ^{Se}G than on T and ^{Se}T.

The bond lengths and bond angles of the planar and buckled ground-state equilibrium geometries of T in the gas phase calculated using MP2/6-311+G(d,p) are shown in Table C5 in Appendix C. The bond lengths and bond angles hardly change between the planar and buckled structures.

The planar (or near-planar) geometries obtained for the solvated molecules, using SCRF-PCM/MP2/6-311+G(d,p) implies that adding solvation with SCRF-PCM removes the intramolecular basis set deficiency of MP2/6-311+G(d,p).

7.3.2.1.5 Effect of Se Substitution on the Ground-state Equilibrium Geometries

From above, it can be seen that B3LYP/6-311+G(d,p) and HF/6-311+G(d,p) levels of theory can successfully predict the torsion angles of the molecules in the gas phase, whereas MP2 cannot. In addition, comparison of the calculated bond lengths and bond angles of T and G in the gas phase using B3LYP/6-311+G(d,p) and the experimental structures (from gas-phase electron diffraction (GED) and microwave (MW) data[53], and high-resolution X-ray and neutron crystallography[54], presented in

Table C6 in Appendix C), shows that the majority of the calculated bond lengths and bond angles are in good agreement. This suggests that B3LYP/6-311+G(d,p) can be used as a reliable method to predict the molecular structure. Therefore this method has been used here to investigate the ground-state structural differences between the natural bases and the Se-base analogues. The bond lengths and bond angles of T, ^{Se}T, G and ^{Se}G are shown in Figure 7.17.

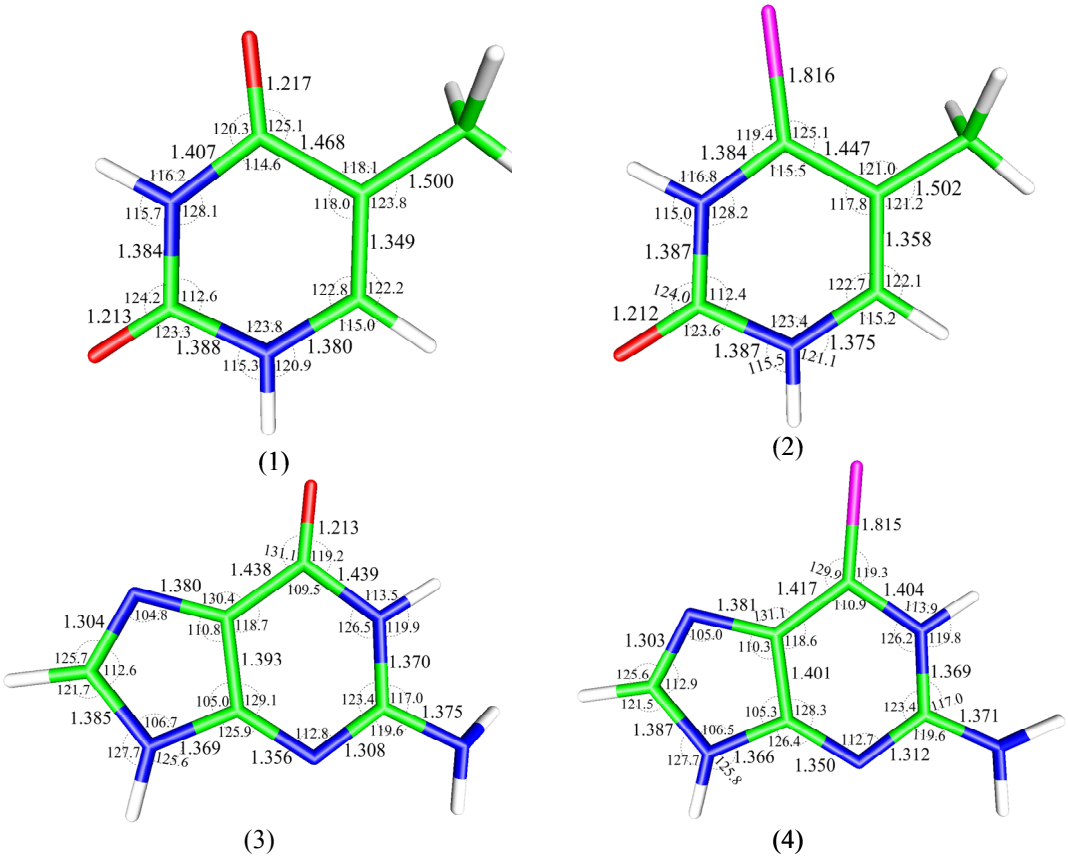


Figure 7.17 The bond lengths and bond angles of the ground-state equilibrium geometire of (1) T, (2) ^{Se}T, (3) G and (4) ^{Se}G optimised using B3LYP/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen and pink for selenium.

Comparing the structures of T and ^{Se}T, all the bond lengths are similar except for the C_4-Se bond of ^{Se}T, which is ~0.6 Å longer than C_4-O in T, and the C_4-C_5 and C_4-N_3 bonds, which are ~0.02 Å shorter in ^{Se}T than in T. These changes of the bond lengths are consistent with those obtained from MP2/6-31G(d,p) level by Vazaquez-Mayagoitia *et al.*[31]. They suggested that the huge increase of C_4-Se bond length is consistent with the van der Waals radius of Se being 0.6 Å larger than O[32]. The differences of the bond angles between T and ^{Se}T are less than 1 ° except for

C_4-C_5-C and C-C_5-C_6 (~3 °). This can also be related to the size of Se. The torsion angles are the same between T and ^{Se}T, both of which are planar. Therefore, replacement of O by Se in T has minimal effect on the molecular geometry in the gas phase, which is consistent with the previous report of Vazaquez-Mayagoitia *et al.*[31]. .

Comparing the structures of G and ^{Se}G, the C_6-Se bond is 0.6 Å longer than C_6-O in G, and C_6-N_1 and C_5-C_6 in ^{Se}G decrease by 0.035 Å and 0.021 Å, respectively. These changes are consistent with those obtained from HF/DZP by Leszcyznski *et al.*[34]. There is little difference in bond angles between G and ^{Se}G, with maximum of 1.4 °. The torsion angles in ^{Se}G are also similar to those in G, which is again consistent with Leszcyznski *et al.*[34]. Therefore, replacing O by Se in G has minimal effect on the gas phase structure.

Comparing the solution-phase structures shows that the differences between T/G and ^{Se}T/ ^{Se}G in solution are similar to those in the gas phase. Therefore, the replacement of O by Se in T and G, in both gas phase and solution, does not significantly change the molecular structure.

7.3.2.1.6 Dipole Moment

The ground-state dipole moments of T, ^{Se}T, G and ^{Se}G at different levels of theory are shown in Table 7.6. The dipole moments of the Se-base analogues are greater than those of the natural bases, irrespective of the level of theory applied. This trend is consistent with that for G and ^{Se}G obtained from HF/DZP[34]. Upon solvation, the dipole moment of each molecule is ~1.5 times that in the gas phase. The dipole moment of T calculated at B3LYP/6-311+G(d,p) level in this work is extremely close to that obtained experimentally by the photodetachment-photoelectron spectrum of the isolated T monomer in the gas phase (4.6 Debye (D))[55]. The value for T calculated at MP2/6-311+G(d,p) level in this work is moderately greater than those obtained at MP2/6-31++G(d,p)[56] and MP2/cc-PVDZ[9] levels, which are 4.60 D and 3.84 D, respectively. The dipole moment of G obtained at B3LYP/6-311+G(d,p) level

in this work is very similar to those obtained from B3LYP/TZVP[57], B3LYP/cc-pVDZ[13] and B3LYP/aug-cc-pVTZ[58], which are 6.78D, 6.45D and 6.71D, respectively. The dipole moments of G and ^{Se}G calculated using HF/6-311+G(d,p) in this work are more than 0.8 D greater than the corresponding ones from HF/DZP[34]. This suggests that the dipole moment obtained from the B3LYP method changes little with different basis sets, whereas values from MP2 and HF methods depend on the basis set used.

Method and basis set	Dipole moment/ D			
	T	^{Se}T	G	^{Se}G
B3LYP/6-311+G(d,p)	4.53	5.16	6.71	7.74
HF/6-311+G(d,p)	4.77	5.95	6.89	8.89
MP2/6-311+G(d,p)	5.01	5.84	7.10	8.44
SCRM-PCM/MP2/6-311+G(d,p)	7.06	9.66	10.4	13.9

Table 7.6 The dipole moments of the optimised ground-state equilibrium geometries of T, ^{Se}T, G and ^{Se}G at different levels of theory.

7.3.2.2 The First Excited-state Equilibrium Geometry

The optimised first excited-state equilibrium geometries of T, ^{Se}T, G and ^{Se}G in the gas phase were calculated using CIS/6-311+G(d,p), based on the ground-state MP2/6-311+G(d,p) optimised geometry, and in solution (water) were calculated using SCRF-PCM/CIS/6-311+G(d,p).

7.3.2.2.1 Bond Lengths and Bond Angles

The calculated bond lengths for T and ^{Se}T, and for G and ^{Se}G are shown in Tables C7 and C8 in Appendix C and the bond angles are shown in Tables C9 and C10 in Appendix C.

The comparisons of the bond lengths and bond angles of G, in both gas phase and solution, between this work and those obtained previously from CIS/cc-pVDZ[13] are

shown in Table C11 in Appendix C. The gas phase values are in moderately good agreement with the largest deviation of bond lengths being 0.06 Å and that of bond angles being 6 °. The values in solution are extremely similar to those from integral equation formalism PCM (IEF-PCM)/CIS/cc-pVDZ[13] with the largest deviation of bond lengths being 0.022 Å and that of bond angles being 1 °. It appears that adding solvation through PCM reduces the difference of bond lengths and angles of G between different basis sets. This suggests that the calculated first excited-state equilibrium geometry of G in solution using SCRF-PCM/CIS/6-311+G(d,p) is reliable.

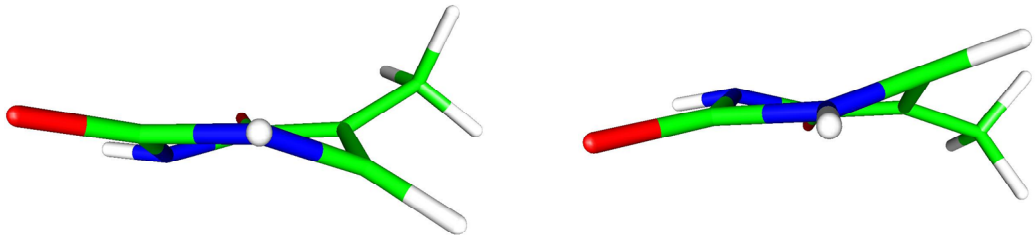
Comparing the first excited-state equilibrium geometries of T and ^{Se}T in the gas phase with those in solution at CIS level (see Tables C7 and C9 in Appendix C), the largest deviation of bond lengths is 0.022 Å and that of bond angles is 2 º. Therefore, solvation does not have significant effect on the bond lengths and bond angles of first excited-state equilibrium geometries of T and ^{Se}T.

The maximum bond length difference for G and ^{Se}G in the gas phase compared with the corresponding ones in solution is ~0.05 Å; and that of bond angle is 7 °. Therefore, solvation has a bigger effect than that for T and ^{Se}T.

7.3.2.2.2 Planarity and Torsion Angles

As shown in Figure 7.18 and Figure C7 in Appendix C, the first excited-state equilibrium geometries of T in both gas phase and solution are buckled with the two O atoms, CH_3 group, N_3 and C_6 atoms out of the plane of the six-membered ring. The non-planar structure in the gas phase is consistent with that obtained from CIS/4-31G[11], TDDFT/PB30/6-31G(d)[12] and CASSCF[9]. The solvated structure is consistent with that of U calculated using PCM/(TD)PBE0/6-31G(d)[12]. There is a noticeable phenomenon that all the signs of the torsion angle values (see Figure C7 in Appendix C) are opposite between the gas phase structure (from now on called T_{gas}) and the solution-phase structure (from now on called T_{sol}), indicating that the

structures of T_{gas} and T_{sol} are almost mirror-images. The torsion angles of T_{sol} are slightly smaller than those of T_{gas}; therefore T_{sol} is slightly less buckled than T_{gas}.



(1) (2)

Figure 7.18 The side view of the first excited-state equilibrium geometry of T (1) in the gas phase at CIS/6-311+G(d,p) level; and (2) in solution at SCRF-PCM/CIS/6-311+G(d,p) level. Green stands for carbon, blue for nitrogen, white for hydrogen, and red for oxygen.

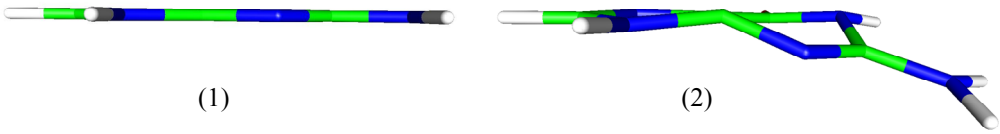
This suggests the existence of two excited-state conformations, with the gas phase and solution-phase calculations having converged to alternative forms. To confirm this, a gas-phase optimisation was performed using T_{sol} as the starting geometry, and a PCM calculation using T_{gas} as the starting geometry. The optimised structures, corresponding to the mirror-images of the structures of T_{sol} and T_{gas} (Figure 7.18) are shown in Figure C8 in Appendix C. The bond lengths and bond angles of the two conformers of T in both gas phase and solution are shown in Table C12 in Appendix C. The two conformers have almost identical geometries except that they are bent to opposite directions and their calculated energies are also almost identical, in both gas phase and solution.

As shown in Figure 7.19 and Figure C9 in Appendix C, the first excited-state equilibrium geometry of ^{Se}T in the gas phase is slightly buckled, with Se out of the plane of the six-membered ring, whereas in solution it is planar.

(1) (2)

Figure 7.19 The side view of the first excited-state equilibrium geometry of ^{Se}T (1) in the gas phase at CIS/HF/6-311+G(d,p) level; and (2) in solution at SCRF-PCM/CIS/6-311+G(d,p) level. Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen and pink for selenium.

As shown in Figure 7.20 and Figure C10 in Appendix C, the first excited-state equilibrium geometry of G in the gas phase is found to be planar. This is different from many previous results, such as those obtained from TDDFT/TZVP[14], CIS/cc-pVDZ[13], CIS/4-31G mixed basis set[8], and CIS/6-311G(d,p)[24]. Moreover, the higher level of theory CASSCF/6-31G(d)[10] also shows that the first excited-state equilibrium geometry of G is non-planar. Upon solvation, the six-membered ring is distorted with N_3, C_2 atoms and the amino group out of the plane, which is consistent with that of G in the gas phase obtained from (TD)B3LYP/TZVP[14], CIS/4-31G mixed basis set[8] and CIS/6-311G(d,p)[24]. However, the out of plane direction for G in solution in this work is opposite to that of G in solution reported previously from PCM/CIS/cc-pVDZ[13] and in the gas phase CASSCF/6-31G(d)[10] (see Section 7.1 Introduction), which could possibly be due to the existence of different conformers, as seen above for T.



(1) (2)

Figure 7.20 The side view of the first excited-state equilibrium geometry of G (1) in the gas phase at CIS/6-311+G(d,p) level; and (2) in solution at SCRF-PCM/CIS/6-311+G(d,p) level. Green stands for carbon, blue for nitrogen, white for hydrogen, and red for oxygen.

As shown in Figure 7.21 and Figure C11 in Appendix C, the first excited-state equilibrium geometry of ^{Se}G in the gas phase and in solution is almost planar. The torsion angles in solution are similar to those in the gas phase.

(1) (2)

Figure 7.21 The side view of the first excited-state equilibrium geometry of ^{Se}G (1) in the gas phase at CIS/6-311+G(d,p) level; and (2) in solution at SCRF-PCM/CIS/6-311+G(d,p) level. Green stands for carbon, blue for nitrogen, white for hydrogen, and pink for selenium.

7.3.2.2.3 Effect of Se Substitution on the First Excited-state Equilibrium Geometries

Comparing the first excited-state equilibrium geometries of T and ^{Se}T in the gas phase, the bond lengths of the six-membered ring change significantly, due to the structure of T being much more buckled than that of ^{Se}T. The C_4-Se bond length is about 0.63 Å longer than C_4-O in T. There is a significant difference of bond angles between T and ^{Se}T for those of which the centre atoms are C_4 and C_6, with a maximum difference up to ~7 °. The torsion angles of T are much bigger than those of ^{Se}T. Therefore, replacement of O by Se has a substantial effect on the first excited-state equilibrium geometry of T in the gas phase.

Comparing the first excited-state equilibrium geometries of T and ^{Se}T in solution, the structural differences are similar to those in the gas phase.

Comparing G and ^{Se}G in the gas phase, there are differences in the bond lengths in the six-membered ring, with the maximum change being roughly 0.06 Å. The C_6-Se bond length is about 0.68 Å longer than C_6-O in G. The differences of bond angles are relatively small with maximum less than 2 °. G is planar, whereas the six-membered ring of ^{Se}G is slightly non-planar (torsion angles less than 4 °) and the five-membered ring is almost planar.

Comparing G and ^{Se}G in solution, C_4-C_5 and C_4-C_9 bond lengths in the five-membered ring of ^{Se}G decrease up to ~0.05 Å. The C_6-Se bond length is about 0.64 Å longer than C_6-O in G. The differences of other bond lengths between G and ^{Se}G are no more than 0.03 Å. C_5-C_6-Se in ^{Se}G reduces by 11 ° relative to C_5-C_6-O in G, whereas C_6-N_1-C_2 and C_5-C_4-N_3 in the six-membered ring increase by up to 6 ° in ^{Se}G. The differences of other bond angles are less than 3 ° between G and ^{Se}G. The structure of G is much more buckled and distorted than that of ^{Se}G. Therefore, the difference of the first excited-state structure between G and ^{Se}G in solution is quite large, particularly with respect to the six-membered ring.

In conclusion, replacement of O by Se in T and G has significant effect on the first excited-state equilibrium structural properties in both gas phase and solution.

7.3.2.2.4 Dipole Moment

The first excited-state dipole moments of T, ^{Se}T, G and ^{Se}G in the gas phase and solution at various levels of theory are shown in Table 7.7. The values, in both gas phase and solution, calculated at (TD)B3LYP/6-311+G(d,p) level are smaller than those obtained at CIS/6-311+G(d,p) level. The dipole moments of ^{Se}T and ^{Se}G in both gas phase and solution are slightly greater than those of T and G, irrespective of the level of theory applied. The dipole moment of G in the gas phase at CIS/6-311+G(d,p) (7.76 D) is much bigger than that obtained from CIS/cc-pVDZ (4.50 D)[13]. However, the dipole moment of G in solution (10.53 D) at SCRF-PCM/CIS/6-311+G(d,p) is reasonably similar to that calculated at IEF-PCM/CIS/cc-pVDZ, 9.95 D[13].

Method and basis set	Dipole moment/ D			
	T	^{Se}T	G	^{Se}G
(TD)B3LYP/6-311+G(d,p)	4.76	5.06	7.31	7.25
SCRF-PCM/(TD)B3LYP/6-311+G(d,p)	5.09	5.10	7.12	8.14
CIS/ 6-311+G(d,p)	5.32	5.86	7.76	8.19
SCRF-PCM/CIS/6-311+G(d,p)	7.92	8.81	10.53	14.04

Table 7.7 The dipole moments of first excited-state equilibrium geometries of T, ^{Se}T, G and ^{Se}G at different levels of theory.

7.3.2.3 Singlet Absorption Transitions

7.3.2.3.1 Transition Wavelengths, Oscillator Strengths and Transition Types

The first fifteen electronic singlet absorption transitions in the gas phase were calculated using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p) and in solution (water) were obtained using SCRF-PCM/CIS/6-311+G(d,p). From now on, the

absorption wavelengths calculated, using CIS/6-311+G(d,p) and SCRF-PCM/CIS/6-311+G(d,p), are presented after scaling.

CIS/6-311+G(d,p) and SCFR/CIS/6-311+G(d,p) for calculations of the absorption spectra in the gas phase and in solution are based on the ground-state MP2/6-311+G(d,p) optimised geometry in the gas phase, which is slightly buckled for the four molecules. To determine whether the slight deviation from planar geometry affects the predicted transition wavelength and oscillator strength, a CIS/6-311+G(d,p) calculation based on the MP2/6-311+G(d,p) planar geometry of T in the gas phase (a saddle point structure with one imaginary frequency (see Section 7.3.2.1.4)) was performed, the results of which are shown in Table C13 in Appendix C. It can be seen that the calculated transitions and their oscillator strengths are hardly changed by the slight difference in planarity of the starting structures. Therefore, the results obtained using CIS/6-311+G(d,p) and SCFR/CIS/6-311+G(d,p) based on the slightly non-planar geometry, shown below, are reliable.

Table 7.8 shows the electronic singlet absorption transitions occurring above 200 nm for T in the gas phase and in solution, along with their oscillator strengths and transition types.

Excited -state	(TD)B3LYP/6-311+G(d,p)			CIS/6-311+G(d,p)			SCRF-PCM/CIS/6-311+G(d,p)		
	Type	λ/ nm	f	Type	λ/ nm	f	Type	λ/ nm	f
1	n-π*	261	0	π-π*	277	0.430	π-π*	283	0.480
2	π-π*	248	0.138	n-π*	275	0.003	n-π*	256	≈0
3	π-σ*	223	≈0	π-σ*	260	0.005	π-σ*	250	0.009
4	n-π*	212	≈0	π-σ*	230	0.008	π-σ*	224	0.001
5	π-π*	207	0.069	n-π*	224	0.003	π-π*	219	0.077
6				π-σ*	223	≈0	π-σ*	219	0.007
7				π-π*	218	0.008	π-π*	217	0.246
8				π-π*	211	0.204	n-π*	212	0.015
9				π-π*	201	0.390	π-σ*	200	0.047

Table 7.8 The singlet absorption transitions occurring over 200 nm (scaled by a factor of 0.72 for CIS method), along with their oscillator strengths and transition types, for T in the gas phase calculated using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p), and in solution using SCRF-PCM/CIS/6-311+G(d,p).

Assignment of the character of the molecular orbitals is based on visual inspection. However, some orbitals were seen to have a mixed character (π, σ and n), one example is illustrated in Figure 7.22. It can be seen than the highest occupied molecular orbital (HOMO) -1 orbital of T at (TD)B3LYP/6-311+G(d,p) level is of n (non-bonding) type at two O atoms and σ type at the six-membered ring. This orbital is classified as n type. This is consistent with the classification of Gustavsson *et al.*[12] and Varsano *et al.*[16].

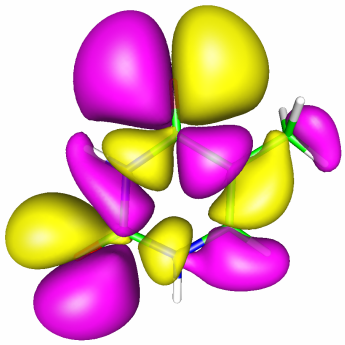


Figure 7.22 The HOMO-1 orbital of T in the gas phase calculated using (TD)B3LYP/6-311+G(d,p). Contour value is 0.01.

The calculated first absorption transition of T in the gas phase obtained using (TD)B3LYP/6-311+G(d,p) occurs at 261 nm (f = 0), and is a forbidden n to π* transition. The second transition occurs at 248 nm (f = 0.138) and is a π to π* transition. The results for the first two transitions are very consistent with those obtained from (TD)B3LYP/6-311++G(d,p)[5] and CASPT2[22]. Another absorption transition with significant oscillator strength occurs at 207 nm (f = 0.069) and is a π to π* transition.

The calculated first absorption transition in the gas phase from CIS/6-311+G(d,p) occurs at 277 nm (f = 0.430), and is a π to π* transition; the second transition occurs at 275 nm (f = 0), a forbidden n to π* transition. These wavelengths are consistent with those obtained from CIS/4-31G[11] with the largest deviation of 16 nm, whereas the transition types and oscillator strengths are interchanged compared with those obtained from CIS/4-31G[11] and the higher level of theory CASPT2[22]. Other

absorption transitions with significant oscillator strength occur at 211 nm (f = 0.204) and 201 nm (f = 0.390), both π to π* transitions.

The calculated first absorption transition of T in solution occurs at 283 nm (f = 0.480) and is a π to π* transition. The second transition occurs at 256 nm (f = 0) and is a forbidden n to π* transition. As in the gas phase, the transition types and oscillator strengths for these transitions are interchanged compared with CASPT2[22]. Another absorption transition with high oscillator strength occurs at 217 nm (f = 0.246) and is also a π to π* transition.

Table 7.9 shows the first fifteen electronic singlet absorption transitions for ^{Se}T in the gas phase and in solution, along with their oscillator strengths and transition types.

Excited -state	(TD)B3LYP/6-311+G(d,p)			CIS/6-311+G(d,p)			SCRF-PCM/CIS/6-311+G(d,p)		
	Type	λ/ nm	f	Type	λ/ nm	f	Type	λ/ nm	f
1	n-π*	522	≈0	n-π*	484	≈0	n-π*	409	≈0
2	π-π*	331	0.282	π-π*	376	0.646	π-π*	371	0.728
3	n-π*	283	0	π-π*	290	0.029	π-π*	291	0.014
4	π-π*	279	0.015	π-σ*	287	0.021	n-σ*	282	0.042
5	n-σ*	265	0.014	n-σ*	280	0.066	π-σ*	278	0.090
6	π-π*	255	0.077	n-σ*	279	0.021	n-σ*	273	0.073
7	π-σ*	252	0.002	π-σ*	268	0.040	π-σ*	269	≈0
8	n-π*	244	0.001	π-σ*	262	0.007	π-σ*	244	0.033
9	n-σ*	239	0.056	n-σ*	244	0.028	n-σ*	240	0.066
10	n-σ*	228	0.002	π-σ*	243	0.009	n-π*	238	0.002
11	π-σ*	224	0.030	π-π*	242	0.058	π-σ*	236	0.009
12	n-σ*	222	0.007	n-π*	249	0.001	π-π*	235	0.296
13	π-σ*	215	0	n-π*	239	0.007	n-π*	232	0.033
14	π-σ*	209	0.003	π-π*	236	0.041	n-π*	230	0.016
15	n-π*	205	0	π-σ*	235	0.010	π-σ*	221	0.002

Table 7.9 The first fifteen singlet absorption transitions (scaled by a factor of 0.72 for CIS method), along with their oscillator strengths and transition types for ^{Se}T in the gas phase calculated using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p), and in solution using SCRF-PCM/CIS/6-311+G(d,p).

The first absorption transition of ^{Se}T in the gas phase, calculated using (TD)B3LYP/6-311+G(d,p), occurs at 522 nm (f = 0) and is a forbidden n to π*

transition. The second transition occurs at 331 nm (f = 0.282) and is a π to π* transition. Other absorption transitions with oscillator strength > 0.05 occur at 255 nm (f = 0.077) and 239 nm (f = 0.056). The calculated first absorption transition of ^{Se}T in the gas phase using CIS/6-311+G(d,p) occurs at 484 nm (f = 0) and is an n to π* transition (forbidden). The second transition occurs at 376 nm (f = 0.646) and is a π to π* transition. Other absorption transitions with oscillator strength > 0.05 occur at 280 nm (f = 0.066) and 241 nm (f = 0.058).

The calculated first absorption transition of ^{Se}T in solution occurs at 409 nm (f = 0) and is a forbidden n to π* transition, as found in the gas phase, except the absorption peak is blue-shifted ~75 nm in solution. The second transition occurs at 371 nm (f = 0.728) and is a π to π* transition. Another transition occurs at 235 nm (f = 0.296). Other calculated transitions with oscillator strength > 0.05 occur at 278 nm (f = 0.090), 273 nm (0.073) and 240 nm (f = 0.066).

Table 7.10 shows the electronic singlet absorption transitions occurring over 200 nm for G in the gas phase and in solution, along with their oscillator strengths and transition types.

Excited -state	(TD)B3LYP/6-311+G(d,p)			CIS/6-311+G(d,p)			SCRF-PCM/CIS/6-311+G(d,p)		
	Type	λ/ nm	f	Type	λ/ nm	f	Type	λ/ nm	f
1	π-σ*	270	0.002	π-σ*	293	0.005	π-π*	288	0.241
2	π-π*	254	0.145	π-π*	289	0.276	π-σ*	285	0.058
3	π-σ*	241	0.003	π-σ*	269	0.021	π-π*	266	0.586
4	π-π*	239	0.256	π-π*	266	0.411	π-σ*	249	0.015
5	n-π*	232	≈0	n-π*	258	0.002	π-σ*	243	0.021
6	π-σ*	217	0.001	π-σ*	250	0.002	π-π*	242	0.052
7	n-π*	212	≈0	π-σ*	244	0.007	n-π*	240	0.002
8	n-π*	209	0.001	π-π*	243	0.088	π-σ*	230	0.005
9	π-σ*	208	≈0	π-σ*	232	0.023	π-π*	229	0.075
10	n-σ*	207	0.013	π-π*	228	0.064	π-σ*	222	0.001
11	π-π*	203	0.009	n-π*	228	0.017	n-π*	217	0.011
12	n-π*	203	0	π-π*	215	0.012	π-σ*	216	0.005
13	-	-	-	π-π*	213	0.005	π-π*	210	0.102
14	-	-	-	π-π*	211	0.010	π-σ*	204	0.004
15	-	-	-	π-σ*	208	0.002	π-π*	203	0.006

Table 7.10 The first fifteen singlet absorption transitions occurring over 200 nm (scaled by a factor of 0.72 for CIS method), along with their oscillator strengths and transition types for G in the gas phase calculated using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p), and in solution using SCRF-PCM/CIS/6-311+G(d,p).

The calculated first absorption transition of G in the gas phase obtained using (TD)B3LYP/6-311+G(d,p) occurs at 270 nm (f = 0.002) and is a π to σ* transition (forbidden). The second transition occurs at 254 nm (f = 0.145) and is a π to π* transition. The latter transition is similar to first transition calculated using (TD)B3LYP/6-311++G[5], (TD)B3LYP/6-311G(d,p)[24] and the higher level of theory CASPT2/6-31G[10]. This suggests that (TD)B3LYP/6-311+G(d,p) generates the wrong result for the first absorption transition type and oscillator strength. The fourth transition occurs at 239 nm (f = 0.256) and is a π to π* transition, which is also similar to third transition calculated using (TD)B3LYP/6-311++G[5].

The calculated first absorption transition of G in the gas phase using CIS/6-311+G(d,p) occurs at 293 nm (f =0.005) and is a π to σ* transition (forbidden). The second transition occurs at 289 nm (f = 0.276) and is a π to π* transition. The latter transition is similar to first transition calculated using CIS/4-31G mixed basis set[8], CIS/6-311G(d,p)[24] and higher level of theory CASPT/6-31G[10]. This suggests

that CIS/6-311+G(d,p) wrongly predicts the first transition type and oscillator strength, which is similar to (TD)B3LYP/6-311+G(d,p). The fourth transition occurs at 266 nm (f = 0.411) and is a π to π* transition, which is similar to the calculated third absorption transition using CIS/4-31G mixed basis set[8] except for a small red shift of 20 nm.

The calculated first absorption transition of G in solution occurs at 288 nm (f = 0.241), and is a π to π* transition (which is different to the forbidden gas-phase transition). This transition is similar to the first transition calculated using SCRF-PCM/CIS/4-31G mixed basis set[8] and CASPT2/6-31G[10]. The third transition occurs at 266 nm (f = 0.586) and is a π to π* transition. Another transition occurs at 210 nm with considerable oscillator strength ~0.1. All the wavelengths of the calculated absorption transitions in solution are blue-shifted slightly compared with those in the gas phase using CIS/6-311+G(d,p).

Table 7.11 shows the first fifteen electronic singlet absorption transitions for ^{Se}G in the gas phase and in solution, along with their oscillator strengths and transition types.

Excited -state	(TD)B3LYP/6-311+G(d,p)			CIS/6-311+G(d,p)			SCRF-PCM/CIS/6-311+G(d,p)		
	Type	λ/ nm	f	Type	λ/ nm	f	Type	λ/ nm	f
1	n-π*	466	0	n-π*	448	0	π-π*	376	0.734
2	π-π*	341	0.264	π-π*	383	0.621	n-π*	369	0
3	n-π*	318	≈0	π-σ*	325	0.025	π-σ*	301	0.074
4	n-σ*	307	0.050	n-σ*	301	0.038	n-σ*	282	0.024
5	π-σ*	305	0.002	π-π*	297	0.235	π-σ*	280	0.050
6	π-π*	303	0.108	π-σ*	292	0.008	π-π*	279	0.026
7	π-σ*	267	0.004	n-σ*	287	0.025	n-σ*	274	0.024
8	n-σ*	266	0.008	π-σ*	281	0.025	π-π*	269	0.403
9	π-π*	254	0.003	π-π*	273	0.040	π-σ*	261	0.028
10	n-π*	248	0	π-π*	269	0.073	π-π*	253	0.012
11	n-σ*	246	0.025	π-σ*	266	0.007	π-σ*	253	0.018
12	π-σ*	245	0.010	π-π*	263	0.051	n-π*	238	0.003
13	n-π*	243	≈0	n-σ*	253	0.003	π-σ*	237	0.005
14	n-σ*	236	0.020	n-π*	252	0.005	n-σ*	237	0.029
15	π-π*	234	0.029	π-σ*	249	0.001	n-π*	236	0.011

Table 7.11 The first fifteen singlet absorption transitions (scaled by a factor of 0.72 for CIS method), along with their oscillator strengths and transition types for ^{Se}G in the gas phase calculated using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p), and in solution using SCRF-PCM/CIS/6-311+G(d,p).

The calculated first absorption transition of ^{Se}G in the gas phase, obtained using (TD)B3LYP/6-311+G(d,p), occurs at 466 nm (f = 0) and is an n to π* transition (forbidden). The second transition occurs at 341 nm (f = 0.264) and is a π to π* transition. Another transition with considerable oscillator strength occurs at 303 nm (f = 0.108) and is a π to π* transition. The calculated first absorption transition of ^{Se}G in the gas phase, obtained using CIS/6-311+G(d,p), occurs at 448 nm (f = 0) and is an n to π* transition (forbidden). The second transition occurs at 381 nm (f = 0.621) and is a π to π* transition. Other transitions with high oscillator strength occur at 297 nm (f = 0.235) and 269 nm (f = 0.073).

The calculated first absorption transition of ^{Se}G in solution occurs at 376 nm (f = 0.734) and is a π to π* transition, which is completely different to that in the gas phase. The second transition occurs at 369 nm (f = 0) and is an n to π* transition (forbidden). It seems that upon solvation, the first two absorption transitions are interchanged compared with those in the gas phase. There are another two π to π*

transitions with considerable oscillator strength occurring at 301 nm (f = 0.074) and 269 nm (f = 0.403).

7.3.2.3.2 Comparison between Calculated and Experimental Results

Table 7.12 shows a summary of the calculated absorption peaks, from different levels of theory, and the difference, Δλ, between the calculated and experimental wavelengths (see Table 7.5 and Tables 7.8 to 7.11). The oscillator strengths are also shown.

Molecule	(TD)B3LYP/gas				CIS/gas				SCRF-PCM/CIS/sol				Exp/sol	
	peak	Δλ	f	rel f	peak	Δλ	f	rel f	peak	Δλ	f	rel f	peak	rel f
T	248	-16	0.14	1	277	13	0.43	1	283	19	0.48	1	264[27]	0.8
	207	1	0.07	0.5	211	5	0.20	0.5	217	11	0.25	0.5	206[27]	1
^{Se}T	331	-37	0.29	1	376	8	0.65	1	371	3	0.73	1	368	0.8
	-	-	-	-	-	-	-	-	-	-	-	-	203	1
G	254	-21	0.15	0.6	289	14	0.28	0.7	288	13	0.24	0.4	275[28]	0.7
	239	-7	0.27	1	266	20	0.41	1	266	20	0.59	1	246[28]	1
^{Se}G	341	-16	0.26	1	383	26	0.62	1	376	19	0.73	1	357	0.5
	-	-	-	-	-	-	-	-	-	-	-	-	213	1

Table 7.12 A summary of the calculated wavelengths (in nm) of the absorption peaks, from different levels of theory, and the difference, Δλ (in nm), between calculated and experimental wavelengths (a negative value indicates the calculated peaks are blue-shifted relative to the experimental ones, a positive value indicates a red shift of calculated relative to experimental). The calculated oscillator strengths, f, are also shown. For each molecule, relative oscillator, rel f, of the listed transitions is given for comparison with the relative values obtained from the experimental absorption spectra.

The calculated absorption peaks in the gas phase using CIS/6-311+G(d,p) are red-shifted between 27 nm and 45 nm compared with those using (TD)B3LYP/6-311+G(d,p), and the corresponding oscillator strengths using CIS/6-311+G(d,p) are ~2 times or more as much as those using (TD)B3LYP/6-311+G(d,p). The transition types of the first two absorption transitions of T are interchanged between these two levels of theory, whereas those of ^{Se}T, G and ^{Se}G are unchanged between these two levels of theory.

The calculated absorption peaks in the gas phase using (TD)B3LYP/6-311+G(d,p) are blue-shifted up to 37 nm compared with those from experiment, whereas those using CIS/6-311+G are red-shifted up to 26 nm compared with those from experiment. Generally speaking, (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p) both generate moderately satisfactory absorption spectra of T, ^{Se}T, G and ^{Se}G compared with the corresponding experimental absorption spectra. Both methods succeed to predict the ratio of oscillator strengths of the absorption peaks of G, but fail to predict that of T. It is notable that all three methods fail to predict the intense short wavelength transition in both ^{Se}T and ^{Se}G.

Including solvation with SCRF-PCM in CIS/6-311+G(d,p) does not change the calculated absorption peaks dramatically with the largest deviation less than 10 nm for all the four molecules compared with those in the gas phase, which is consistent with the results of Gustavsson *et al.*[12] and Shukla *et al.*[8]. However, solvation has different effect on the calculated absorption peaks for different molecules; the calculated absorption peaks of T in solution are red-shifted, those of G in solution stay almost the same, and the calculated absorption peak of ^{Se}T and ^{Se}G in solution is blue-shifted compared with the corresponding one in the gas phase. All the oscillator strengths are slightly greater than those in the gas phase.

The calculated absorption peaks in solution are red-shifted up to 20 nm compared with the experiment, suggesting SCRF-PCM/CIS generates satisfactory absorption spectra. Including solvation for T does not improve accuracy between the calculated and experimental spectrum, whereas for G the accuracy between the calculated spectrum in solution and experiment is similar to that between the calculated spectrum in the gas phase and experiment. Including solvation for ^{Se}T and ^{Se}G does improve the agreement between the calculated spectrum in solution and experiment. In addition, including solvation reduces the accuracy of predicting the ratio of oscillator strengths of the absorption peaks of G, and also fails to predict that of T.

Table 7.13 shows a comparison of the calculated and experimental red shifts between the longest wavelength absorption band of each Se-base analogue and its

corresponding natural base (see Table 7.5 and Tables 7.8 to 7.11). It can be seen that the calculated red shift is between 83 nm to 99 nm for (TD)B3LYP, CIS and SCRF-PCM/CIS methods with 6-311+G(d,p) basis set, which is consistent with the experimental values (more than 80 nm). Therefore, all of the three methods successfully predict the large red shift in the absorption spectra of T and G, when O is replaced by Se.

Molecule	(TD)B3LYP/gas/ nm	CIS/gas/ nm	SCRF-PCM/CIS/sol/ nm	Exp/sol/ nm
^{Se}T-T	83	99	88	104
^{Se}G-G	87	94	88	82

Table 7.13 A summary of the calculated and experimental red shift between the absorption spectra of the Se-base analogues and the natural bases.

7.3.2.3.3 Effect of Se on the Calculated Absorption Transition

The origin of the effect of Se on the longest wavelength absorption transition is now considered, by examining more closely the results of the SCRF-PCM/CIS/6-311+G(d,p) calculations.

Figure 7.23 shows the major orbitals involved in the longest wavelength intense absorption transitions of T and ^{Se}T, corresponding to the experimental peaks at 264 nm and 368 nm, respectively. For T, the HOMO to LUMO+3 and HOMO to LUMO +4 transitions contribute, whereas for ^{Se}T, the HOMO to LUMO transition predominates. The HOMO orbital of ^{Se}T resembles that of T, but the shape of the LUMO of ^{Se}T is considerably different to the LUMO+3 and LUMO+4 of T. In addition, the oscillator strength of transition in ^{Se}T is 0.25 greater than that in T. Therefore, the red-shifted absorption band in ^{Se}T seems to be due to the appearance of a new transition of different character from the longest wavelength transition in T.

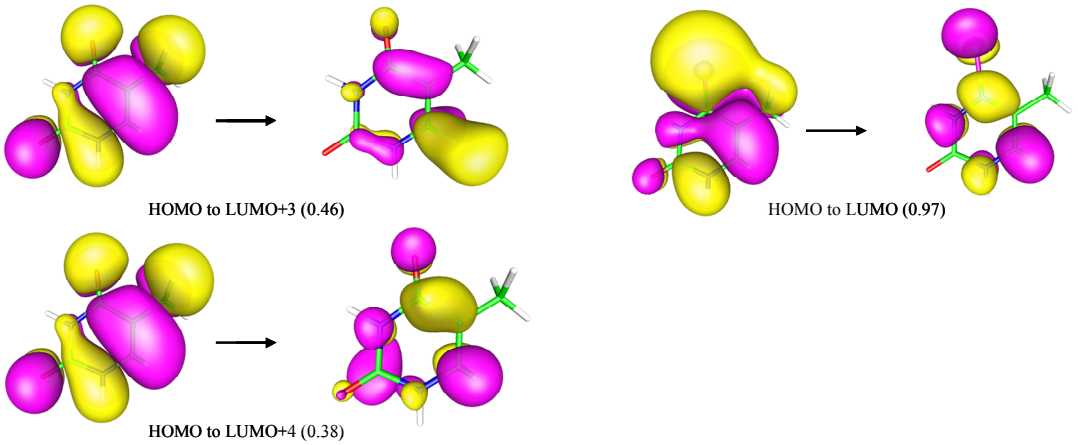


Figure 7.23 The major orbitals involved in the longest wavelength intense absorption transition of (1) T, and (2) ^{Se}T in solution (SCRF-PCM/CIS/6-311+G(d,p)). The contour value for the HOMO is 0.01, and that for unoccupied orbitals is 0.03. The value in the brackets shows the fractional contribution to the transition (obtained from the square of the CI expansion coefficient.) Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen and pink for selenium.

Table 7.14 shows the eigenvalues of the orbitals of T and ^{Se}T presented in Figure 7.23. It can be seen that Se-induced ~90 nm absorption peak shift can be traced to increases in the eigenvalue of HOMO and decreases in the eigenvalues of unoccupied orbitals involved in the transition, and is due predominantly the lower eigenvalue of LUMO of ^{Se}T compared with that of the LUMO+3 and LUMO+4 of T.

T		^{Se}T	
Orbital	**Eigenvalue/ Hartree**	**Orbital**	**Eigenvalue/ Hartree**
HOMO	-0.34	HOMO	-0.32
-	-	LUMO	0.04
LUMO+3	0.08	-	-
LUMO+4	0.10	-	-

Table 7.14 The eigenvalues of the orbitals of T and ^{Se}T presented in Figure 7.23.

Figure 7.24 shows the major orbitals involved in the longest wavelength absorption transitions of G and ^{Se}G, corresponding to the experimental bands at 275 nm and 357 nm, respectively. The orbital shapes of the HOMO and LUMO+8 of ^{Se}G are similar to the corresponding ones of G. Therefore, the HOMO to LUMO+8 transition of ^{Se}G is similar to that of G. In addition, the orbital shape of LUMO+1 of ^{Se}G is similar to

that of LUMO+6 of G, suggesting that the HOMO to LUMO+1 transition of ^{Se}G is similar to HOMO to LUMO+6 transition of G. Although, the oscillator strength of the first absorption transition of ^{Se}G is ~0.5 bigger than that of G, the first transition of ^{Se}G appears to be similar in character to the transition of G. Thus the effect of Se on the absorption spectrum of G seems to be due to a red-shift of the longest wavelength transition of G, rather than a completely new transition.

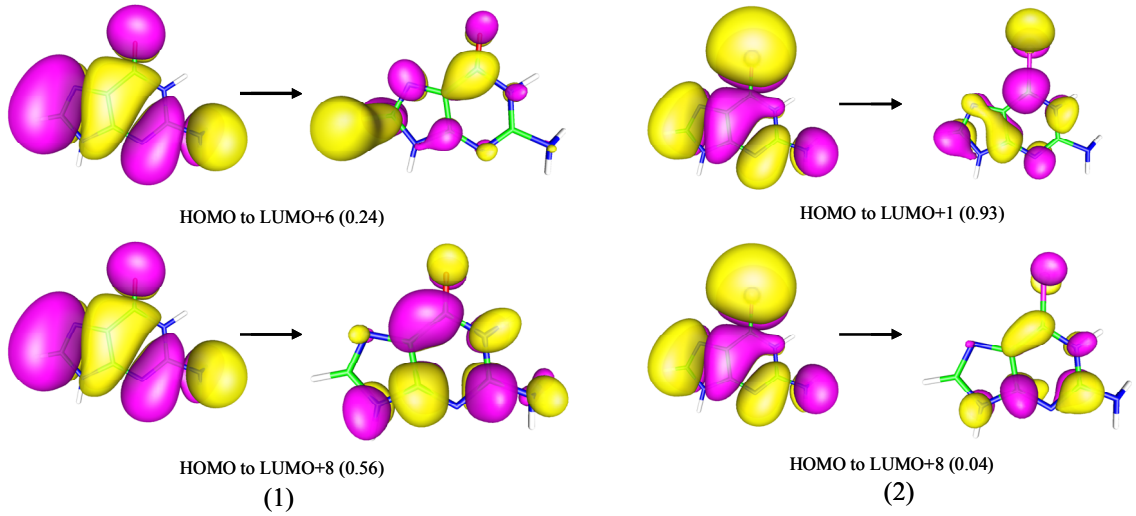


Figure 7.24 The major orbitals involved in the longest wavelength absorption transition of (1) G, and (2) ^{Se}G in solution (SCRF-PCM/CIS/6-311+G(d,p)). The contour value for HOMO is 0.01, and that for unoccupied orbitals is 0.03. The value in the brackets shows the fractional contribution to the transition (obtained from the square of the CI expansion coefficient.) Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen and pink for selenium.

Table 7.15 shows the eigenvalues of the orbitals of G and ^{Se}G presented in Figure 7.24. Similar to ^{Se}T, the Se-induced ~90 nm red-shift in ^{Se}G can be traced to increases in the eigenvalue of HOMO and decreases in the eigenvalues of unoccupied orbitals involved in the transition, predominantly the eigenvalue decrease between the LUMO +1 of ^{Se}G and the LUMO+6 of G (0.06).

G		^{Se}G	
Orbital	**Eigenvalue/ Hartree**	**Orbital**	**Eigenvalue/ Hartree**
HOMO	-0.31	HOMO	-0.30
LUMO+6	0.11	LUMO+1	0.05
LUMO+8	0.13	LUMO+8	0.11

Table 7.15 The eigenvalues of the orbitals of G and ^{Se}G presented in Figure 7.24. The orbitals in the same row have similar shape between G and ^{Se}G.

7.3.2.4 Singlet Emission Transitions

7.3.2.4.1 Transition Wavelengths, Oscillator Strengths and Transition Types

The first three singlet emission transitions in the gas phase were calculated using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p), and in solution were calculated using SCRF-PCM/(TD)B3LYP/6-311+G(d,p) and SCRF-PCM/CIS/6-311+G(d,p). From now on, the wavelengths of the transitions calculated using CIS/6-311+G(d,p) and SCRF-PCM/CIS/6-311+G(d,p) are presented after scaling.

Table 7.16 shows the first three singlet emission transitions of T, along with their oscillator strengths and transition types. As noted above, in the excited-state, T exists in two conformations, in the gas phase and solution. The emission transitions calculated on the basis of one of the conformational geometries are given in Table 7.16. The calculations for the other conformer are shown in Table C14 in Appendix C. It can be seen that the emission transitions are almost identical between the two conformers.

Excited-state	(TD)B3LYP/6-311+G(d,p)			SCRF-PCM/(TD)B3LYP/6-311+G(d,p)			CIS/6-311+G(d,p)			SCRF-PCM/CIS/6-311+G(d,p)		
	Type	**λ/ nm**	**f**	**Type**	**λ/ nm**	**f**	**Type**	**λ/ nm**	**f**	**Type**	**λ/ nm**	**f**
1	π-π*	379	0.086	π-π*	346	0.112	π-π*	435	0.214	π-π*	398	0.292
2	n-π*	318	0.008	n-π*	321	0.001	π-π*	298	0.026	π-π*	276	0.017
3	n-π*	261	0.007	n-π*	260	0.013	n-π*	280	0.007	n-π*	259	0.004

Table 7.16 The calculated first three singlet emission transitions for T in the gas phase using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p) and in solution using SCRF-PCM/(TD)B3LYP/6-311+G(d,p) and SCRF-PCM/CIS/6-311+G(d,p) (scaled by a factor of 0.72 for CIS method), along with their oscillator strengths and transition types.

The first emission transition of T in the gas phase, calculated using (TD)B3LYP/6-311+G(d,p), occurs at 379 nm (f = 0.086) and is a π to π* transition. The first emission transition in solution, calculated using SCRF-PCM/(TD)B3LYP/6-311+G(d,p), occurs at 346 nm, a π to π* transition with oscillator strength slightly greater than that in the gas phase. This is similar to the emission wavelength (353 nm) obtained using PCM/(TD)PB30/6-311+G(d,p)[12], whereas the oscillator strength is roughly half as much as that obtained using PCM/(TD)PB30/6-311+G(d,p)[12].

The first emission transition of T in the gas phase, calculated using CIS/6-311+G(d,p), occurs at 435 nm (f = 0.214) and is a π to π* transition. This wavelength is significantly longer (> 100 nm) than that from CIS/4-31G[11]. The first emission transition in solution obtained from SCRF-PCM/CIS/6-311+G(d,p) occurs at 398 nm, a π to π* transition with oscillator strength slightly greater than that in the gas phase.

Table 7.17 shows the first three singlet emission transitions of ^{Se}T, along with their oscillator strengths and transition types.

Excited-state	(TD)B3LYP/6-311+G(d,p)			SCRF-PCM/(TD)B3LYP/6-311+G(d,p)			CIS/6-311+G(d,p)			SCRF-PCM/CIS/6-311+G(d,p)		
	Type	**λ/ nm**	**f**	**Type**	**λ/ nm**	**f**	**Type**	**λ/ nm**	**f**	**Type**	**λ/ nm**	**f**
1	n-π*	574	≈0	n-π*	535	≈0	n-π*	532	≈0	n-π*	421	≈0
2	π-π*	344	0.269	π-π*	327	0.294	π-π*	400	0.530	π-π*	375	0.705
3	π-π*	290	0.010	π-π*	286	0.009	π-π*	294	0.083	π-π*	290	0.024

Table 7.17 The calculated first three singlet emission transitions for ^{Se}T in the gas phase using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p) and in solution using SCRF-PCM/(TD)B3LYP/6-311+G(d,p) and SCRF-PCM/CIS/6-311+G(d,p) (scaled by a factor of 0.72 for CIS method), along with their oscillator strengths and transition types.

The first emission transition of ^{Se}T in the gas phase is predicted by (TD)B3LYP/6-311+G(d,p) to occur at 574 nm and is an n to π* transition with f = 0. Upon solvation, the wavelength decreases to 535 nm, with oscillator strength and transition type in solution same as in the gas phase.

The first emission transition of ^{Se}T in the gas phase predicted by CIS/6-311+G(d,p) occurs at 532 nm and is a forbidden n to π* transition. Upon solvation, the wavelength shifts to 421 nm, with oscillator strength and transition type unchanged.

Table 7.18 shows the calculated first three singlet emission transitions of G, along with their oscillator strengths and transition types.

Excited-state	(TD)B3LYP/6-311+G(d,p)			SCRF-PCM/(TD)B3LYP/6-311+G(d,p)			CIS/6-311+G(d,p)			SCRF-PCM/CIS/6-311+G(d,p)		
	Type	**λ/ nm**	**f**	**Type**	**λ/ nm**	**f**	**Type**	**λ/ nm**	**f**	**Type**	**λ/ nm**	**f**
1	π-σ*	310	0.002	π-π*	330	0.106	π-σ*	343	≈0	π-π*	384	0.202
2	π-σ*	270	0.005	π-π*	299	0.007	π-σ*	314	0.007	π-σ*	304	0.062
3	π-π*	269	0.233	π-σ*	265	0.001	π-π*	305	0.388	π-σ*	288	0.399

Table 7.18 The calculated first three singlet emission transitions for G in the gas phase using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p) and in solution using SCRF-PCM/(TD)B3LYP/6-311+G(d,p) and SCRF-PCM/CIS/6-311+G(d,p) (scaled by a factor of 0.72 for CIS method), along with their oscillator strengths and transition types.

The first emission transition of G in the gas phase, calculated using (TD)B3LYP/6-311+G(d,p), occurs at 310 nm and is a π to σ* transition (forbidden). This wavelength is consistent with that obtained using (TD)B3LYP/TZPP[14] and CASPT2/6-31G[10], whereas it is considerably shorter (~150 nm) than that obtained using (TD)B3LYP/6-311G(d,p)[24]. The latter large discrepancy is probably due to the different basis set used. The transition type is different to the π to π* transition predicted by (TD)B3LYP/TZPP[14], (TD)B3LYP/6-311G(d,p)[24] and CASPT2/6-31G[10] .

The calculated first transition in solution (SCRF-PCM/(TD)B3LYP/6-311+G(d,p)) occurs at 330 nm (f = 0.106) and is a π to π* transition, in contrast to the n to π* transition predicted in the gas phase

The first emission transition in the gas phase predicted by CIS/6-311+G(d,p) occurs at 343 nm (f = 0); this is blue-shifted ~50 nm compared with CIS/6-311G(d,p)[24] and is considerably shorter (~130 nm) than that from CIS/4-31G mixed basis set[8]. Again, the large discrepancy is possibly due to the different basis set used. The oscillator strength indicates this calculated transition is forbidden, which is different to CIS/6-311G(d,p)[24] and CIS/4-31G mixed basis set[8]. This transition is a π to σ* transition, which is different to CIS/6-311G(d,p)[24] and CASPT2/6-31G[10] (π to π* transition).

In solution (SCRF-PCM/CIS/6-311+G(d,p)) the transition occurs at 384 nm (f = 0.202) and is a π to π* transition, in contrast to the gas phase. The wavelength is significantly shorter (~70 nm) than that from SCRF-PCM/CIS/4-31G mixed basis set[8], whereas the oscillator strength is similar to SCRF-PCM/CIS/4-31G mixed basis set[8].

Table 7.19 shows the calculated first three singlet emission transitions of ^{Se}G, along with their oscillator strengths and transition types.

Excited-state	(TD)B3LYP/6-311+G(d,p)			SCRF-PCM/(TD)B3LYP/6-311+G(d,p)			CIS/6-311+G(d,p)			SCRF-PCM/CIS/6-311+G(d,p)		
	Type	λ/ nm	f	Type	λ/ nm	f	Type	λ/ nm	f	Type	λ/ nm	f
1	n-π*	617	≈0	π-π*	629	≈0	n-π*	575	≈0	π-π*	484	0.303
2	π-π*	402	0.200	π-π*	443	0.171	π-π*	467	0.370	n-π*	441	0.081
3	n-π*	318	0.024	π-σ*	337	0.004	π-π*	324	0.037	π-π*	322	0.069

Table 7.19 The calculated first three singlet emission transitions for ^{Se}G in the gas phase using (TD)B3LYP/6-311+G(d,p) and CIS/6-311+G(d,p) and in solution using SCRF-PCM/(TD)B3LYP/6-311+G(d,p) and SCRF-PCM/CIS/6-311+G(d,p) (scaled by a factor of 0.72 for CIS method), along with their oscillator strengths and transition types.

The first emission transition of ^{Se}G in the gas phase, calculated using (TD)B3LYP/6-311+G(d,p), occurs at 617 nm (f = 0) and is an n to π* transition (forbidden). The calculated first emission transition of ^{Se}G in solution using SCRF-PCM/(TD)B3LYP/6-311+G(d,p) occurs at 629 nm (f = 0) and is a forbidden π to π* transition.

The first emission transition in the gas phase, calculated using CIS/6-311+G(d,p), occurs at 575 nm (f = 0) and is a forbidden n to π* transition. In solution (SCRF-PCM/CIS/6-311+G(d,p)) the transition shifts to 484 nm and becomes a π to π* transition, with f = 0.303.

7.3.2.4.2 Comparison between Calculated and Experimental Results

Table 7.20 shows a summary of the calculated emission wavelengths, from different levels of theory, the difference, Δλ, between the calculated and experimental emission wavelengths, the oscillator strengths, and the experimental quantum yields.

Molecule	(TD)B3LYP/gas			SCRF-PCM/(TD)B3LYP/sol			CIS/gas			SCRF-PCM/CIS/sol			Exp/sol	
	peak	Δλ	f	peak	Δλ	f	peak	Δλ	f	peak	Δλ	f	peak	quantum yield
T	379	50	0.086	346	17	0.112	435	106	0.214	398	69	0.292	329[12]	1.2×10^{-4} [29]
^{Se}T	574	-	0	535	-	0	532	-	0	421	-	0	-	-
G	310	-25	0.002	330	-5	0.106	343	8	0	384	49	0.202	335[28]	0.8×10^{-4} [29]
^{Se}G	617	167	0	629	179	0	575	125	0	484	34	0.303	450	1.2×10^{-2}

Table 7.20 A summary of the calculated emission wavelengths (in nm), from different levels of theory, the difference, Δλ, between the calculated and experimental emission wavelengths, the oscillator strengths, and the experimental quantum yields.

The CIS/6-311+G(d,p)-calculated emission wavelengths of T and G, in both gas phase and solution, are red-shifted more than 30 nm compared with those from (TD)B3LYP/6-311+G(d,p). The CIS/6-311+G(d,p)-calculated oscillator strengths are ~2 times as much as those from (TD)B3LYP/6-311+G(d,p). The emission wavelengths of ^{Se}T and ^{Se}G predicted by CIS/6-311+G(d,p) are blue-shifted more than 40 nm in the gas phase and more than 110 nm in solution, compared with those from (TD)B3LYP/6-311+G(d,p). With the exception of the high oscillator strength for ^{Se}G in solution calculated using SCRF-PCM/CIS/6-311+G(d,p), the oscillator strengths of ^{Se}T and ^{Se}G in both gas phase and solution are predited to be 0 at both levels of theory.

Including solvation in (TD)B3LYP/6-311+G(d,p), shifts the emission peaks of T and ^{Se}T to the blue by up to 39 nm, whereas those of G and ^{Se}G are red-shifted up to 20 nm. Including solvation in CIS/6-311+G(d,p) shifts the emission peak of T to the blue by ~40 nm; those of ^{Se}T and ^{Se}G are also blue-shifted, but by than 90 nm, whereas the emission of is G red-shifted by about 50 nm. Broadly speaking, including solvation in both levels of theory increases the oscillator strengths. In general the discrepancy in emission wavelength between calculation and experiment is reduced by inclusion of solvation in the calculation. This is consistent with the findings of Gustavsson *et al*.[12]

SCRF-PCM/(TD)B3LYP/6-311+G(d,p) predicts the emission peak of T and G satisfactorily with the largest deviation between the calculated and experimental peak less than 17 nm. SCRF-PCM/CIS/6-311+G(d,p) gives moderately satisfactory prediction for T and G, with the calculated peak red-shifted more than 49 nm. Although, both levels of theory predict considerable oscillator strength of T and G in solution (> 0.1), the experimental quantum yields of T and G are extremely low[29]. As described in the introduction, this is because of very rapid, barrierless, non-radiative decay via conical intersections.

Both levels of theory predict the emission oscillator strength of ^{Se}T to be zero, which is consistent with the experimental observation that ^{Se}T*p is not fluorescent. This suggests that the lowest singlet excited-state of ^{Se}T does not necessarily undergo fast non-radiative decay.

SCRF-PCM/(TD)B3LYP/6-311+G(d,p) fails to predict the wavelength of the emission peak of ^{Se}G observed experimentally at ~450 nm. In addition, the calculated oscillator strength of this transition is 0, which is not consistent with the observation that ^{Se}G*p is weakly fluorescent. However, SCRF-PCM/CIS/6-311+G(d,p) predicts an emission wavelength for ^{Se}G in reasonable agreement with experiment, with a deviation of 34 nm. The oscillator strength is predicted to be high which would suggest that the low quantum yield of ^{Se}G*p (1.2 x 10^{-2}) is due to rapid non-radiative decay, perhaps by a similar mechanism to that found in G.

7.3.2.4.3 Effect of Se on the Calculated Emission Transition

Figure 7.25 shows the major orbitals involved in the first transition of T and ^{Se}T, calculated by SCRF-PCM/CIS/6-311+G(d,p). The difference in the nature of the orbitals is clearly apparent. The oscillator strength of the ^{Se}T nπ* is 0, whereas that of the T ππ*is high. Therefore, the first transition of ^{Se}T is a new transition.

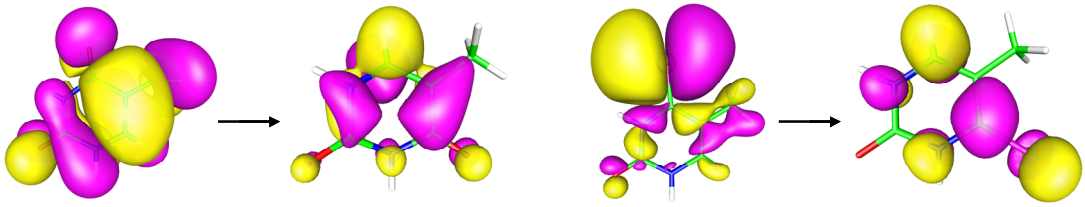


Figure 7.25 The major orbitals involved in the first transition of (1) T and (2) ^{Se}T in solution (SCRF-PCM/CIS/6-311+G(d,p)). The contour value for occupied orbitals is 0.01, and that for unoccupied orbitals is 0.035. The value in the brackets shows the fractional contribution to the transition (obtained from the square of CI expansion coefficient) Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen and pink for selenium.

Figure 7.26 shows the major orbitals involved in the first transition of G and ^{Se}G, calculated by SCRF-PCM/CIS/6-311+G(d,p). The HOMO to LUMO transition of

^{Se}G is similar to the HOMO to LUMO+4 transition of G, and the HOMO to LUMO+1 transition of ^{Se}G is similar to the HOMO to LUMO transition of G. In addition, the oscillator strength of the first transition of ^{Se}G is only roughly 0.1 bigger than G. Therefore, the first transition of ^{Se}G resembles that of G, but is shifted to lower energy.

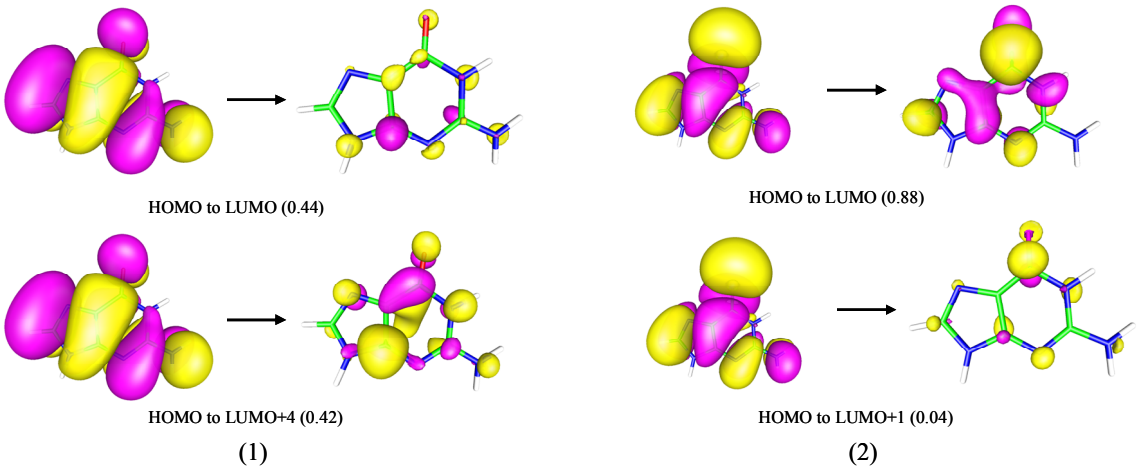


Figure 7.26 The major orbitals involved in the first transition of (1) G and (2) ^{Se}G in solution (SCRF-PCM/CIS/6-311+G(d,p)). The contour value for occupied orbitals is 0.01, and that for unoccupied orbitals is 0.035. The value in brackets shows the fracional contribution to the transition (obtained from the square of CI expansion coefficient). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen and pink for selenium.

Table 7.21 shows the eigenvalues of the orbitals of G and ^{Se}G presented in Figure 7.26. The eigenvalue of HOMO of ^{Se}G is the same as for G, and that of LUMO+1 is also the same to that of LUMO of G. Therefore, the Se-induced 100 nm calculated emission peak shift in ^{Se}G is due to the significant decrease in the eigenvalue of the LUMO of ^{Se}G compared with that of the LUMO+4 of G.

G		^{Se}G	
Orbital	**Eigenvalue/ Hartree**	**Orbital**	**Eigenvalue/ Hartree**
HOMO	-0.28	HOMO	-0.28
LUMO	0.06	LUMO+1	0.06
LUMO+4	0.09	LUMO	0.04

Table 7.21 The eigenvalues of the orbitals of G and ^{Se}G presented in Figure 7.26. The orbitals in the same row have similar shape between G and ^{Se}G.

7.4 Conclusion

Replacement of O by Se in T and G causes almost no change in ground-state equilibrium geometries, in both the gas phase and solvated aqueous state. The only significant change in bond length is an increase in the C-Se bond lengths (C_4-Se of ^{Se}T and C_6-Se of ^{Se}G) which increase about 0.6 Å due to the increases van der Waals radius of Se. As for the natural bases, the Se-base analogues are planar.

In contrast, replacement of O by Se has a significant effect on the first excited-state equilibrium geometry of both T and G. The first excited-state of T is predicted to have a significantly buckled equilibrium structure, with the largest torsion angle of $C_4C_5C_6N_1$ being ~37 °, whereas the first excited-state equilibrium structure of ^{Se}T is slightly buckled in the gas phase, with Se out of the six-membered ring, and planar in solution. The first excited-state equilibrium geometry of G in solution is non-planar, whereas that of ^{Se}G in solution is almost planar.

The longest wavelength absorption bands of ^{Se}T*p and ^{Se}G*p are observed experimentally at 368 nm and 357 nm, respectively, red-shifted by more than 80 nm relative to the natural bases. The singlet absorption transitions predicted using SCRF-PCM/CIS/6-311+G(d,p) are in good agreement with experiment for all four molecules, T, ^{Se}T, G and ^{Se}G, such that the Se-induced red-shift is successfully predicted. The calculations show that the red-shift of the ^{Se}T absorption spectrum is due to the appearance of a new long-wavelength transition, whereas in ^{Se}G it is due to a reduction in energy of the corresponding transition of G. The Se-induced absorption peak shift can be traced predominantly to decreases in the eigenvalues of unoccupied orbitals involved in the transitions of the Se-substituted bases.

^{Se}T*p is not fluorescent, while ^{Se}G*p is weakly fluorescent, with an emission maximum at 450 nm, which is ~120 nm longer than that of G. The low quantum yield of ^{Se}G*p means that it is not a promising fluorescent base analogue. SCRF-PCM/CIS/6-311+G(d,p) is moderately successful at predicting the emission wavelength of ^{Se}G, and the magnitude of the red-shift relative to G. The calculation

also shows that the emission transition of ^{Se}G is due to red-shifting the corresponding transition of G and is not a new transition. The Se-induced red-shift is due to the decrease in the eigenvalue of unoccupied orbital involved in the transition. It is evident that the lowest singlet $\pi\pi^*$ excited-state of ^{Se}G experiences a rapid non-radiative decay pathway, as does the natural base.

7.5 References

(1) Caton-Williams, J.; Huang, Z. *Angew. Chem. Int. Edit.* **2008**, *47*, 1723-1725.

(2) Salon, J.; Jiang, J.; Sheng, J.; Gerlits, O. O.; Huang, Z. *Nucleic Acids Res.* **2008**, *36*, 7009-7018.

(3) Salon, J.; Sheng, J.; Jiang, J.; Chen, G.; Caton-Williams, J.; Huang, Z. *J. Am. Chem. Soc.* **2007**, *129*, 4862-4863.

(4) Cavaluzzi, M. J.; Borer, P. N. *Nucleic Acids Res.* **2004**, *32*, e13-22.

(5) Shukla, M. K.; Leszczynski, J. *J. Comput. Chem.* **2004**, *25*, 768-778.

(6) Czader, A.; Bittner, E. R. *J. Chem. Phys.* **2008**, *128*, 035101.

(7) Kushwaha, P. S.; Kumar, A.; Mishra, P. C. *Spectrochim. Acta A* **2004**, *60*, 719-728.

(8) Shukla, M. K.; Mishra, S. K.; Kumar, A.; Mishra, P. C. *J. Comput. Chem.* **2000**, *21*, 826-846.

(9) Perun, S.; Sobolewski, A. L.; Domcke, W. *J. Phys. Chem. A* **2006**, *110*, 13238-13244.

(10) Chen, H.; Li, S. *J. Phys. Chem. A* **2006**, *110*, 12360-12362.

(11) Shukla, M. K.; Mishra, P. C. *Chem. Phys.* **1999**, *240*, 319-329.

(12) Gustavsson, T.; Banyasz, A.; Lazzarotto, E.; Markovitsi, D.; Scalmani, G.; Frisch, M. J.; Barone, V.; Improta, R. *J. Am. Chem. Soc.* **2006**, *128*, 607-619.

(13) Mennucci, B.; Toniolo, A.; Tomasi, J. *J. Phys. Chem. A* **2001**, *105*, 7126-7134.

(14) Marian, C. M. *J. Phys. Chem. A* **2007**, *111*, 1545-1553.

(15) Chen, H.; Li, S. *J. Chem. Phys.* **2006**, *124*, 154315.

(16) Varsano, D.; Di Felice, R.; Marques, M. A. L.; Rubio, A. *J. Phys. Chem. B* **2006**, *110*, 7129-7138.

(17) Tsolakidis, A.; Kaxiras, E. *J. Phys. Chem. A* **2005**, *109*, 2373-2380.

(18) Silaghi, S. D.; Friedrich, M.; Scholz, R.; Kampen, T. U.; Cobet, C.; Esser, N.; Richter, W.; Braun, W.; Zahn, D. R. T. *Thin Solid Films* **2004**, *455-456*, 505-508.

(19) Yamazaki, S.; Domcke, W.; Sobolewski, A. L. *J. Phys.Chem. A* **2008**, *112*, 11965-11968.

(20) Serrano-Andres, L.; Merchan, M.; Borin, A. C. *J. Am. Chem. Soc.* **2008**, *130*, 2473-2484.

(21) Fulscher, M. P.; Serrano-Andres, L.; Roos, B. O. *J. Am. Chem. Soc.* **1997**, *119*, 6168-6176.

(22) Lorentzon, J.; Fuelscher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, *117*, 9265-9273.

(23) Zechmann, G.; Barbatti, M. *J. Phys. Chem. A* **2008**, *112*, 8273-8279.

(24) Zhang, L.; Bu, Y. *J. Phys. Chem. B* **2008**, *112*, 10723-10731.

(25) Ozeki, K.; Sakabe, N.; Tanaka, J. *Acta Crystallogr. B* **1969**, *25*, 1038-1045.

(26) Thewalt, U.; Bugg, C. E.; Marsh, R. E. *Acta Crystallogr. B* **1971**, *27*, 2358-2363.

(27) Anex, B. G.; Fucaloro, A. F.; Dutta-Ahmed, A. *J. Phys. Chem.* **1975**, *79*, 2636-2639.

(28) Fujiwara, T.; Kamoshida, Y.; Morita, R.; Yamashita, M. *J. Photoch. Photobio. B* **1997**, *41*, 114-121.

(29) Peon, J.; Zewail, A. H. *Chem. Phys. Lett.* **2001**, *348*, 255-262.

(30) Santhosh, C.; Mishra, P. C. *J. Mol. Struct.* **1989**, *198*, 327-337.

(31) Vazquez-Mayagoitia, A.; Huertas, O.; Brancolini, G.; Migliore, A.; Sumpter, B. G.; Orozco, M.; Luque, F. J.; Di Felice, R.; Fuentes-Cabrera, M. *J. Phys. Chem. B* **2009**, *113*, 14465-14472.

(32) Veazey, W. R.; Hodgman, C. D. *Handbook of Chemistry and Physics*; Chemical Rubber Pub. Co.: 1980.

(33) Mautner, H. G.; Chu, S.; Jaffe, J. J.; Sartorelli, A. C. *J. Med. Chem.* **1963**, *6*, 36-39.

(34) Leszczynski, J. *J. Mol. Struct.* **1994**, *311*, 37-44.

(35) Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem.* **1969**, *244*, 1228-1237.

(36) Jiao, D.; Wang. H.; Zhang, Y.; Tang, Y. *Can. J. Chem.* **2009**, *87*, 406-415.

(37) Elshakre, M. *Int. J. Quantum Chem.* **2005**, *104*, 1-15.

(38) Brown, R. D.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Chem. Soc. Chem. Comm.* **1989**, 37-38.

(39) Jang, Y. H.; Goddard, W. A.; Noyes, K. T.; Sowers, L. C.; Hwang, S.; Chung, D. S. *J. Phys. Chem. B* **2002**, *107*, 344-357.

(40) Rejnek, J.; Hanus, M.; Kabelac, M.; Ryjacek, F.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2006-2017.

(41) Yi, H.; Li, X.; Yang, S.; Duan, X. *Int. J. Quantum Chem.* **2003**, *94*, 23-35.

(42) Longworth, J. W.; Rahn, R. O.; Shulman, R. G. *J. Chem. Phys.* **1966**, *45*, 2930-2939.

(43) Gerdil, R. *Acta Crystallogr.* **1961**, *14*, 333-344.

(44) Asturiol, D.; Duran, M.; Salvador, P. *J. Chem. Theory Comput.* **2009**, *5*, 2574-2581.

(45) Torii, H.; Ishikawa, A.; Takashima, R.; Tasumi, M. *J. Mol. Struc.: THEOCHEM* **2000**, *500*, 311-321.

(46) Asturiol, D.; Duran, M.; Salvador, P. *J. Chem. Phys.* **2008**, *128*, 144108.

(47) Roman, M. B. *J. Chem. Phys.* **2008**, *129*, 164101.

(48) Hobza, P.; Havlas, Z. *Theor. Chem. Accounts: Theor. Comput. Model. (Theor. Chim. Acta)* **1998**, *99*, 372-377.

(49) Salvador, P.; Paizs, B.; Duran, M.; Suhai, S. *J. Comput. Chem.* **2001**, *22*, 765-786.

(50) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553-566.

(51) Salvador, P.; Duran, M.; Dannenberg, J. J. *J. Phys. Chem. A* **2002**, *106*, 6883-6889.

(52) Hugas, D.; Simon, S.; Duran, M. *Chem. Phys. Lett.* **2004**, *386*, 373-376.

(53) Vogt, N.; Khaikin, L. S.; Grikina, O. E.; Rykov, A. N.; Vogt, J. *J. Phys. Chem. A* **2008**, *112*, 7662-7670.

(54) Clowney, L.; Jain, S. C.; Srinivasan, A. R.; Westbrook, J.; Olson, W. K.; Berman, H. M. *J. Am. Chem. Soc.* **1996**, *118*, 509-518.

(55) Schiedt, J.; Weinkauf, R.; Neumark, D. M.; Schlag, E. W. *Chem. Phys.* **1998**, *239*, 511-524.

(56) Svozil, D.; Frigato, T.; Havlas, Z.; Jungwirth, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 840-845.

(57) Jones, D. B.; Wang, F.; Winkler, D. A.; Brunger, M. J. *Biophys. Chem.* **2006**, *121*, 105-120.

(58) Zhu, Q.; Wang, F.; Ivanova, E. P. *J. Synchrotron Radiat.* **2009**, *16*, 545-552.

Chapter 8

Conclusions

The aims of this thesis were to investigate base dynamics and base stacking interactions in DNA single strands and dinucleotides; monitor the interactions between DNA and a polymerase Pfu-Pol, using the steady-state and time-resolved fluorescence of 2-aminopurine (2AP); and explore a new class of base analogues in which oxygen is replaced with selenium, specifically at the 4 position of thymine and the 6 position of guanine, using a combination of experiment and quantum chemical computation.

8.1 DNA Single Strands and Dinucleotides

As in DNA duplexes, 2AP in both single strands and dinucleotides shows a long wavelength emission band, with maximum intensity at about 450 nm, when excited at around 360 nm, in addition to the 'normal' emission at 370 nm, which is excited at 310 nm. This long wavelength emission is due to a ground-state heterodimer species formed between 2AP and one of its neighbouring bases. Although, 2AP can form the heterodimer species with either the 5' or 3' neighbouring base, it appears to preferentially form the dimer with the natural base on its 5' side. Dimer formation is less favourable in dinucleotides than in single strands, in which dimer formation is in turn less favourable than in DNA duplexes. The 2AP heterodimer has its own excitation spectrum, implying that in single strands and dinucleotides it is a ground-state species, similar to that seen in DNA duplexes.

The influence of base stacking interaction on DNA conformation and the mechanism of inter-base quenching of 2AP were investigated by measuring the time-resolved fluorescence of 2AP-containing dinucleotides. The fluorescence decay behaviour of

2AP in dinucleotides in PBS buffer at room temperature is similar to that in single strands and described by 4 discrete components, indicating comparable conformational heterogeneity in both cases. The shortest fluorescence lifetime of 2AP in dinucleotides varies with the identity of the neighbouring base: when the neighbouring base is adenine (A) or inosine (I), the shortest lifetime component is absent; when it is guanine (G) or thymine (T), the shortest lifetime is ~60 ps; with cytosine (C), the shortest lifetime is 160 ps. The magnitude of the shortest lifetime component correlates with the driving force for electron transfer (ET) quenching, suggesting ET is the dominant mechanism of quenching of 2AP* in dinucleotides. Broadly speaking, each of the other three lifetimes τ_2, τ_3 and τ_4 have similar values, ~0.7 ns, ~2.3 ns and ~9 ns, for all the five dinucleotides. The dominant component changes from τ_1 in 2AP-G to τ_2 in 2AP-A and τ_3 in other three dinucleotides. Less than 6% of the 2AP population, in all the dinucleotides, has a fluorescence lifetime about 9 ns, corresponding to a fully unstacked conformation.

2AP-A appears to be anomalous in lacking the shortest lifetime component, but this is also the case for the APA single strand. This suggests the fast ET does not occur between 2AP excited-state (2AP*) and A in either of these systems. It is proposed that 2AP and A cannot form the necessary π-stacking interaction to facilitate the fast ET quenching. The different π-stacking interaction behaviour of 2AP-A to the other dinucleotides might be due to the structural similarity between 2AP and A: the exocyclic amine group, present in both 2AP and A, might hinder the formation of a highly π-stacked conformation.

The π-stacking interaction and the imperfectly π-stacked conformations in dinucleotides appears similar to those in single strands. However, the internal base motion and dynamic disorder in dinucleotides is greater than in single strands. The population of highly stacked states decreases significantly and transfers to the weakly stacked states in dinucleotides compared with single strands. The multiplicity of base π-stacking interactions decreases the internal base motion and limits the conformational fluctuations of the single strand structure, as manifested in the fluorescence behaviour of 2AP.

The importance of thermal fluctuations of the bases in facilitating inter-base quenching was studied by measuring the fluorescence lifetimes of 2AP-containing dinucleotides in LiCl at both room temperature and 77 K, the methodology applied previously to 2AP-labelled duplexes. However, unlike the duplexes, the fluorescence response of 2AP-containing dinucleotides is dramatically changed by dissolution in LiCl at room temperature, such that they all show a dominant, short lifetime of ~60 ps. The LiCl solvent appears to enhance the inter-base interactions, giving rise to a short-lived excited-state with emission that is slightly red-shifted compared with the 'normal' 2AP excited-state. The highly concentrated LiCl makes the local environment of the dinucleotides become hydrophobic, which encourages the adoption of highly stacked conformations and enhances inter-base interaction.

The fluorescence response of 2AP-containing dinucleotides is significantly affected by freezing to 77 K, such that the first two shortest decay components vanish. Freezing eliminates the base motion, and as a consequence blocks the fast ET quenching. The frozen dinucleotides show a delocalised conformation formed between 2AP and its neighbouring base with an increased lifetime ~14 ns, which is significantly longer than that of the free 2AP riboside under the same conditions. The majority of the conformational population, in which 2AP was highly quenched at room temperature, exists in the unquenched delocalized conformation at 77 K.

The effect of hydrogen bonding on the conformational changes and inter-bases interactions in DNA was further elucidated through measuring the fluorescence decay behaviour of 2AP in single strands. In general, the fluorescence response of 2AP at 370 nm (the short wavelength emission band) in single strands is quite similar to that in double strands and is described by four discrete components, indicative of comparable conformational heterogeneity in both cases. (The APA sequence context is anomalous as noted above). The major conformation accounts for over 40% of the 2AP population in single strands and has a short lifetime of less than 120 ps. A small fraction, less than 10%, of the 2AP population has a lifetime of around 7 ns. Broadly speaking, the magnitudes of τ_1 and τ_2 increase and the magnitudes of τ_3 and τ_4 decrease in single strands compared with those in DNA duplexes. The population of

the shortest lifetime component decreases significantly and mainly transfers to the second and third components in single strands compared with double strands, resulting in the total population of the second and third components being larger than that of shortest lifetime component. Therefore, less quenched (less stacked) conformations are more highly populated in single strands. Thus, hydrogen bonding increases the extent of the π-stacking interaction, decreases the internal base motion and restricts fluctuations in the DNA duplex structure. This has an important effect on the populations of the various conformational states.

The fluorescence decay of the 2AP dimer (~450 nm emission band) in single strands has four lifetimes, indicating the heterogeneity of the inter-base interactions of this species, which is similar to the behaviour seen in the DNA duplex. The magnitudes of τ_2, τ_3 and τ_4 increase in single strands compared to those in double strands. The absence of hydrogen bonding has a similar effect on the fluorescence decay behaviour of the 2AP dimer and the 2AP monomer in single strands. All the fluorescence lifetimes of the 2AP dimer are at least twice those of 2AP monomer, in single strands. Compared to the 2AP monomer, the 2AP dimer is geometrically constrained and π-stacking interaction between the 2AP dimer and its neighbours is weaker in single strands, resulting in slower quenching rates. The population differences between different conformational states are smaller for the dimer than the monomer implying that this too is influenced by π-stacking interaction.

In order to understand more completely the effects of π-stacking interaction and hydrogen bonding, a number of future studies can be suggested. It would be instructive to investigate the time-resolved fluorescence of 2AP in different lengths of DNA single strands, such as 5 bases or 50 bases long, to compare with the ~20-base strands and dinucleotides studied in the present work. Secondly, the time-resolved fluorescence of 2AP located at different positions in given single strands, for example moving 2AP gradually from 5' side to 3' side, could be measured. Thirdly, the time-resolved fluorescence of 2AP in single strands and their complementary DNA duplexes could be studied together. In order to further and more accurately investigate the role of thermal fluctuations of the bases in facilitating

inter-base quenching, solvents other than 10 M LiCl, which are suitable for freezing but do not have a profound effect on the fluorescence decay of 2AP in dinucleotides, could be sought.

8.2 DNA Unwinding by a DNA Polymerase

The hypothesis that the proof-reading mechanism of archeal family-B polymerase involves unwinding the primer-template duplex region, upon encountering the misincorpated uracil (U), was investigated by time-resolved fluorescence measurements. The 2AP fluorescence decay was used successfully to investigate the structural distortion of primer-templates, with U at the +2 position (U+2) or +4 position (U+4) from the replication fork, by an archaeal family-B DNA polymerase which lacks exonuclease activity (exo$^-$). The fluorescence decay of 2AP in the U+2-exo$^-$ complex shows a dramatic increase in the lifetime of the shortest decay component, and provides convincing evidence for a decrease in the amount of double stranded DNA on polymerase exo$^-$ binding. The fluorescence decay behaviour of 2AP in U+2-exo$^-$ complex is entirely consistent with unwinding of the duplex, leaving 2AP in a poorly stacked, unpaired state. In contrast, a much smaller perturbation of the 2AP fluorescence decay is seen in the U+4-exo$^-$ complex, indicating less profound DNA distortion. It is concluded that the presence of U at the +2 position leads to more DNA unwinding than observed with U+4, but even with U+4 some unravelling may occur. The structural distortions inferred here are consistent with the crystal structural studies of the primer-templates-polymerase complex, and support the hypothesis that the mechanism of archeal family-B polymerase involves unwinding of the primer-template duplex region upon encountering the misincorporated U.

2AP fluorescence also successfully reported the reaction process of the exonuclease of the wild type polymerase (WT), excising the 3' terminus primer bases in primer-templates, albeit with limited time resolution. The speed of excising the first two bases from 3' terminus of primer in U+2 is much faster than in U+4. In addition,

the speed of excising bases in the 3' terminus of primer in Mg^{2+} buffer is significantly faster than in Ca^{2+} buffer. These results are consistent with those from a previous polymerase excision gel assay.

8.3 Se-containing Base Analogues

In recent years, a new class of base analogues in which selenium (Se) replaces oxygen (O) at the 4-position of thymine (^{Se}T) or at 6-position of guanine (^{Se}G) bases has been synthesised. These new base analogues are of potential interest as spectroscopic probes because they show absorption peaks that are greatly red-shifted (by ~100 nm) compared with the natural bases, taking their absorption spectra into the visible region. In order to examine whether ^{Se}T and ^{Se}G can be used as fluorescent base analogues, the UV absorption spectra, fluorescence and phosphorescence emission spectra of 4-Se-thymidine-3'-phosphate (^{Se}T*p) and 6-Se-2'-deoxyguanosine-3'-phosphate (^{Se}G*p) have been studied. To investigate structural difference between the Se-base analogues and the corresponding natural bases, and the origins of the different photophysical properties, the base analogues, ^{Se}T, and ^{Se}G, and the corresponding natural bases, T and G, were studied by quantum chemical calculations.

Quantum chemical computation has shown that replacement of O by Se leaves the optimised equilibrium ground-state geometries of ^{Se}T and ^{Se}G, in both gas phase and solution (water), almost unchanged compared with the corresponding T and G. The only significant structural difference is an increase of about 0.6 Å in the C-Se bond length. However, the Se substituent has a significant influence on the first excited-state structures of the bases, causing a change from buckled to nearly planar.

SCRF-PCM/CIS/6-311+G(d,p) calculations have successfully predicted the absorption spectra of T, ^{Se}T, G and ^{Se}G and reproduced the large red-shifts induced by replacement of O with Se. In the case of ^{Se}T, the shift can be attributed to the appearance of a new transition, whereas in ^{Se}G, the effect appears to be due to the

decrease in energy of a G-like transition. In both cases, the decrease in absorption energy appears to be due mainly to decreases in the eigenvalues of unoccupied orbitals involved in the transitions.

The Se-containing base analogues have no potential value as fluorescent probes: ^{Se}T*p is non-fluorescent and ^{Se}G*p shows only very weak fluorescence. SCRF-PCM/CIS/6-311+G(d,p) calculations predict the oscillator strength of the ^{Se}T fluorescence transition to be zero, consistent with the experimentally observed absence of fluorescence. However, ^{Se}G is predicted to show a high oscillator-strength, $\pi\pi^*$ fluorescence transition, implying that the weak fluorescence observed experimentally is due to efficient non-radiative decay, perhaps by a similar mechanism (rapid, barrierless internal conversion) to that found in G.

Since the quantum chemical calculations can successfully predict the Se-induced absorption and emission peak shift, they offer a promising approach to the future investigation of the photophysical properties of other potential Se-derivatised nucleotides prior to experimentally synthesising and studying these molecules.

Appendix A

Fit Parameters for Fluorescence Decay Curves

Global lifetimes/ ns: τ_2=0.68, τ_3=2.4, τ_4=9.6. Global χ^2=1.057

Emission wavelength/ nm	A_2	A_3	A_4	Local χ^2
370	0.31	0.54	0.15	1.016
390	0.32	0.53	0.15	1.090
400	0.34	0.51	0.15	1.064

Table A1 The global decay parameters for single strand APA excited at 309 nm.

Global lifetimes/ ns: τ_1=0.06, τ_2=0.59, τ_3=2.4, τ_4=6.9. Global χ^2=1.051

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
370	0.51	0.19	0.22	0.08	1.012
385	0.50	0.19	0.22	0.09	1.071
400	0.51	0.20	0.21	0.08	1.070

Table A2 The global decay parameters for single strand GPA excited at 309 nm.

Global lifetimes/ ns: τ_1=0.11, τ_2=0.52, τ_3=2.1, τ_4=7.2. Global χ^2=1.067

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
370	0.44	0.38	0.13	0.05	1.056
385	0.42	0.39	0.13	0.06	1.051
400	0.44	0.37	0.13	0.66	1.096

Table A3 The global decay parameters for single strand APT excited at 309 nm.

Global lifetimes/ ns: τ_1=0.05, τ_2=0.53, τ_3=2.7, τ_4=7.5. Global χ^2=1.046

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
370	0.69	0.11	0.10	0.10	1.086
385	0.69	0.11	0.10	0.10	0.997
400	0.70	0.12	0.09	0.09	1.055

Table A4 The global decay parameters for single strand GPT excited at 309 nm.

Global lifetimes/ ns: τ_1=0.12, τ_2=0.55, τ_3=1.7, τ_4=5.5. Global χ^2=1.048

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
370	0.41	0.34	0.21	0.04	1.047
385	0.41	0.35	0.20	0.04	1.030
400	0.43	0.34	0.19	0.04	1.068

Table A5 The global decay parameters for single strand CPA excited at 309 nm.

Global lifetimes/ ns: τ_1=0.08, τ_2=0.59, τ_3=2.4, τ_4=6.9. Global χ^2=1.058

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
370	0.47	0.14	0.32	0.07	1.023
385	0.45	0.16	0.31	0.08	1.053
400	0.47	0.17	0.29	0.07	1.098

Table A6 The global decay parameters for single strand GPC excited at 309 nm.

Global lifetimes/ ns: τ_1=0.12, τ_2=0.59, τ_3=3.6, τ_4=14.1. Global χ^2=1.034

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
450	0.23	0.57	0.11	0.09	1.046
470	0.24	0.52	0.13	0.11	1.022

Table A7 The global decay parameters for single strand APA excited at 360nm.

Global lifetimes/ ns: τ_1=0.15, τ_2=1.1, τ_3=4.3, τ_4=15.1. Global χ^2=1.065

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
450	0.39	0.24	0.14	0.23	1.067
470	0.39	0.23	0.14	0.24	1.062

Table A8 The global decay parameters for single strand GPA excited at 360 nm.

Global lifetimes/ ns: τ_1=0.22, τ_2=2.2, τ_3=7.2, τ_4=16.9. Global χ^2=1.050

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
450	0.25	0.19	0.46	0.10	1.074
470	0.25	0.22	0.42	0.11	1.026

Table A9 The global decay parameters for single strand GPT excited at 360 nm.

Global lifetimes/ ns: τ_1=0.18, τ_2=1.2, τ_3=4.5, τ_4=14.2. Global χ^2=1.030

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
450	0.32	0.26	0.25	0.17	1.039
470	0.34	0.25	0.24	0.17	1.021

Table A10 The global decay parameters for single strand CPA excited at 360 nm.

Global lifetimes/ ns: τ_1=0.14, τ_2=1.4, τ_3=5.7, τ_4=15.0. Global χ^2=1.084

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
450	0.31	0.22	0.24	0.23	1.113
470	0.33	0.21	0.23	0.23	1.055

Table A11 The global decay parameters for single strand GPC excited at 360 nm.

Global lifetimes/ ns: τ_2=0.61, τ_3=2.0, τ_4=9.6. Global χ^2=1.035

Emission wavelength/ nm	A_2	A_3	A_4	Local χ^2
360	0.64	0.33	0.03	1.018
380	0.64	0.33	0.03	1.059
400	0.63	0.34	0.03	1.027

Table A13 The global decay parameters for dinucleotide 2AP-A in PBS buffer excited at 305 nm.

Global lifetimes/ ns: τ_1=0.16, τ_2=0.78, τ_3=2.4, τ_4=9.2. Global χ^2=1.034

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
360	0.18	0.12	0.69	0.01	1.052
380	0.17	0.14	0.68	0.01	1.008
400	0.18	0.13	0.68	0.01	1.041

Table A14 The global decay parameters for dinucleotide 2AP-C in PBS buffer excited at 305 nm.

Global lifetimes/ ns: τ_1=0.06, τ_2=0.63, τ_3=1.8, τ_4=10.3. Global χ^2=1.022

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
360	0.50	0.12	0.33	0.05	1.033
380	0.47	0.14	0.33	0.06	1.010
400	0.46	0.14	0.34	0.06	1.024

Table A15 The global decay parameters for dinucleotide 2AP-G in PBS buffer excited at 305 nm.

Global lifetimes/ ns: τ_1=0.07, τ_2=0.65, τ_3=2.3, τ_4=10.2. Global χ^2=1.011

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
360	0.28	0.11	0.59	0.02	0.995
380	0.27	0.11	0.59	0.03	1.020
400	0.28	0.11	0.59	0.02	1.019

Table A16 The global decay parameters for dinucleotide 2AP-T in PBS buffer excited at 305 nm.

Global lifetimes/ ns: τ_2=0.83, τ_3=3.1, τ_4=8.5. Global χ^2=1.053

Emission wavelength/ nm	A_2	A_3	A_4	Local χ^2
360	0.33	0.61	0.06	1.037
380	0.33	0.61	0.06	1.078
400	0.34	0.61	0.05	1.044

Table A17 The global decay parameters for dinucleotide 2AP-I in PBS buffer excited at 305 nm.

Global lifetimes/ ns: τ_1=0.06, τ_2=0.84, τ_3=3.8, τ_4=7.5. Global χ^2=1.059

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
360	0.55	0.16	0.20	0.09	1.066
380	0.63	0.13	0.17	0.07	1.036
400	0.68	0.12	0.14	0.06	1.076

Table A18 The global decay parameters for dinucleotide 2AP-A in LiCl at room temperature excited at 305 nm. Decays are fitted with 4 components.

Global lifetimes/ ns: τ_2=0.10, τ_3=2.0, τ_4=6.3. Global χ^2=1.970

Emission wavelength/ nm	A_2	A_3	A_4	Local χ^2
360	0.54	0.25	0.21	1.998
380	0.61	0.21	0.18	1.919
400	0.66	0.18	0.15	1.994

Table A19 The global decay parameters for dinucleotide 2AP-A LiCl at room temperature excited at 305 nm. Decays are fitted with 3 components.

Global lifetimes/ ns: τ_1=0.06, τ_2=0.69, τ_3=3.9, τ_4=7.0. Global χ^2=1.054

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
360	0.59	0.06	0.30	0.05	1.026
380	0.69	0.05	0.22	0.04	1.060
400	0.75	0.05	0.17	0.03	1.065

Table A20 The global decay parameters for dinucleotide 2AP-C in LiCl at room temperature excited at 305 nm.

Global lifetimes/ ns: τ_1=0.07, τ_2=0.59, τ_3=3.7, τ_4=7.7. Global χ^2=1.054

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
360	0.67	0.10	0.15	0.08	1.028
380	0.71	0.09	0.13	0.07	1.060
400	0.74	0.08	0.11	0.06	1.075

Table A21 The global decay parameters for dinucleotide 2AP-G in LiCl at room temperature excited at 305 nm.

Global lifetimes/ ns: τ_1=0.05, τ_2=0.60, τ_3=3.7, τ_4=5.7. Global χ^2=1.053

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
360	0.64	0.07	0.17	0.12	1.060
380	0.70	0.06	0.14	0.10	1.025
400	0.75	0.06	0.11	0.08	1.076

Table A22 The global decay parameters for dinucleotide 2AP-T in LiCl at room temperature excited at 305 nm.

Global lifetimes/ ns: τ_1=0.06, τ_2=0.96, τ_3=3.2, τ_4=6.0. Global χ^2=1.012

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
360	0.37	0.18	0.22	0.23	1.033
380	0.43	0.17	0.20	0.20	1.054
400	0.51	0.15	0.17	0.17	1.063

Table A23 The global decay parameters for dinucleotide 2AP-I in LiCl at room temperature excited at 305 nm. Decays are fitted with 4 components.

Global lifetimes/ ns: τ_2=0.09, τ_3=1.7, τ_4=5.5. Global χ^2=1.379

Emission wavelength/ nm	A_2	A_3	A_4	Local χ^2
360	0.34	0.31	0.36	1.413
380	0.40	0.29	0.31	1.343
400	0.47	0.26	0.28	1.381

Table A24 The global decay parameters for dinucleotide 2AP-I LiCl at room temperature excited at 305 nm. Decays are fitted with 3 components.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	Φ from spectrum	Φ from fit
2AP-A	0.06	0.60	3.5	6.6	0.82	0.05	0.06	0.06	0.042	0.09
2AP-C	0.05	0.40	3.5	5.7	0.86	0.03	0.05	0.06	0.032	0.08
2AP-G	0.05	0.28	3.0	6.8	0.79	0.12	0.05	0.04	0.024	0.07
2AP-T	0.05	0.46	2.5	5.5	0.82	0.04	0.03	0.02	0.025	0.10
2AP-I	0.05	1.0	3.3	6.0	0.52	0.15	0.15	0.17	0.053	0.23

Table A25 The global fluorescence decay parameters for dinucleotides in LiCl at room temperature which are excited at 305 nm and collected at 35 minutes. A factors are for 380 nm emission. The relative quantum yields calculated from both spectra and fluorescence fit parameters are also shown.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	Φ from spectrum	Φ from fit
2AP-A	0.06	0.82	3.8	7.2	0.73	0.08	0.12	0.06	0.076	0.13
2AP-C	0.05	0.63	3.9	6.4	0.79	0.03	0.13	0.04	0.083	0.11
2AP-G	0.06	0.40	3.5	7.2	0.77	0.09	0.08	0.06	0.046	0.11
2AP-T	0.05	0.59	3.3	5.4	0.77	0.05	0.06	0.12	0.072	0.12
2AP-I	0.06	0.96	3.1	5.9	0.46	0.16	0.18	0.20	0.071	0.26

Table A26 The global fluorescence decay parameters for dinucleotides in LiCl at room temperature which are excited at 305 nm and collected at 70 minutes. A factors are for 380 nm emission. The relative quantum yields calculated from both spectra and fluorescence fit parameters are also shown.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	Φ from spectrum	Φ from fit
2AP-A	0.06	0.81	3.7	7.2	0.66	0.12	0.15	0.07	0.093	0.16
2AP-C	0.05	0.62	3.9	6.7	0.73	0.04	0.18	0.04	0.096	0.14
2AP-G	0.06	0.56	3.7	7.6	0.73	0.08	0.12	0.07	0.064	0.14
2AP-T	0.05	0.56	3.4	5.4	0.73	0.05	0.09	0.13	0.084	0.14
2AP-I	0.06	0.91	3.0	5.9	0.43	0.17	0.20	0.20	0.095	0.27

Table A27 The global fluorescence decay parameters for dinucleotides in LiCl at room temperature which are excited at 305 nm and collected at 104 minutes. A factors are for 380 nm emission. The relative quantum yields calculated from both spectra and fluorescence fit parameters are also shown.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4	Φ from spectrum	Φ from fit
2AP-A	0.06	0.81	3.6	7.3	0.63	0.13	0.16	0.08	0.11	0.17
2AP-C	0.06	0.67	3.9	6.7	0.72	0.05	0.19	0.05	0.11	0.15
2AP-G	0.07	0.61	3.7	7.7	0.72	0.08	0.13	0.07	0.07	0.15
2AP-T	0.05	0.63	3.7	5.6	0.71	0.06	0.12	0.11	0.10	0.15
2AP-I	0.06	0.96	3.2	6.0	0.43	0.17	0.20	0.20	0.11	0.27

Table A28 The global fluorescence decay parameters for dinucleotides in LiCl at room temperature which are excited at 305 nm and collected at 140 minutes. A factors are for 380 nm emission. The relative quantum yields calculated from both spectra and fluorescence fit parameters are also shown.

Global lifetimes/ ns: τ_2=1.3, τ_3=11.4, τ4=16.4. Global χ^2=1.081

Emission wavelength/ nm	A_2	A_3	A_4	Local χ^2
360	0.05	0.54	0.41	1.067
380	0.05	0.45	0.50	1.103
400	0.07	0.40	0.53	1.074

Table A29 The global decay parameters for dinucleotide 2AP-A in LiCl at 77 K excited at 305 nm. Decays are fitted with 3 components.

Global lifetimes/ ns: τ_3=7.6, τ_4=15.0. Global χ^2=1.384

Emission wavelength/ nm	A_2	A_3	Local χ^2
360	0.25	0.75	1.371
380	0.18	0.82	1.299
400	0.16	0.84	1.482

Table A30 The global decay parameters for dinucleotide 2AP-A in LiCl at 77 K excited at 305 nm. Decays are fitted with 2 components.

Emission wavelength/ nm	$\tau_2(A_2)$	$\tau_3(A_3)$	$\tau_4(A_4)$	Local χ^2
360	1.2(0.05)	9.3(0.25)	15.0(0.70)	1.066
380	1.2(0.05)	9.5(0.21)	15.2(0.74)	1.095
400	0.9(0.06)	8.3(0.16)	15.2(0.78)	1.075

Table A31 The single decay parameters for dinucleotide 2AP-A in LiCl at 77 K excited at 305 nm at emission wavelength 360 nm, 380 nm and 400 nm, respectively. Decays are fitted with 3 components.

Emission wavelength/ nm	$\tau_3(A_3)$	$\tau_4(A_4)$	Local χ^2
360	4.5(0.11)	14.1(0.89)	1.155
380	3.9(0.09)	14.4(0.91)	1.167
400	3.5(0.10)	14.6(0.90)	1.198

Table A32 The single decay parameters for dinucleotide 2AP-A in LiCl at 77 K excited at 305 nm at emission wavelength 360 nm, 380 nm and 400 nm, respectively. Decays are fitted with 2 components.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
2AP-A	-	0.65	2.0	9.7	-	0.60	0.33	0.07
2AP-C	-	0.47	2.3	9.0	-	0.19	0.75	0.06
2AP-G	0.07	0.60	1.8	10.3	0.37	0.16	0.35	0.12
2AP-T	0.04	0.66	2.3	9.3	0.22	0.11	0.63	0.04

Table A33 The global decay parameters for dinucleotides in PBS buffer at room temperature excited at 292 nm. A factors are for emission wavelength 380 nm.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
2AP-A	0.33	0.80	2.3	9.5	0.34	0.52	0.12	0.02
2AP-C	0.07	0.47	2.4	8.9	0.36	0.14	0.35	0.15
2AP-G	0.05	0.45	1.8	10.3	0.70	0.09	0.16	0.05
2AP-T	0.05	0.48	2.3	9.2	0.62	0.08	0.28	0.02

Table A34 The global decay parameters for dinucleotides in PBS buffer at room temperature excited at 276 nm. A factors are for emission wavelength 380 nm.

Dinucleotide	τ_1/ ns	τ_2/ ns	τ_3/ ns	τ_4/ ns	A_1	A_2	A_3	A_4
2AP-A	0.42	0.90	2.6	9.5	0.46	0.44	0.08	0.02
2AP-C	0.08	0.52	2.4	9.3	0.39	0.15	0.26	0.20
2AP-G	0.04	0.38	1.8	10.8	0.80	0.08	0.09	0.03
2AP-T	0.04	0.42	2.3	8.8	0.70	0.09	0.20	0.01

Table A35 The single decay parameters for dinucleotides in PBS buffer at room temperature excited at 261 nm and collected at emission wavelength 380 nm.

Global lifetimes/ ns: τ_1=0.04, τ_2=0.64, τ_3=2.8, τ_4=8.1. Global χ^2=1.073

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.80	0.11	0.07	0.02	1.069
380	0.74	0.14	0.09	0.03	1.032
395	0.73	0.14	0.09	0.04	1.119

Table A36 The global decay parameters for U+2 DNA duplex in Ca^{2+} buffer at room temperature excited at 315 nm.

Global lifetimes/ ns: τ_1=0.05, τ_2=0.63, τ_3=2.7, τ_4=8.0. Global χ^2=1.074

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.77	0.12	0.08	0.03	1.076
380	0.71	0.15	0.10	0.04	1.076
395	0.69	0.16	0.11	0.04	1.070

Table A37 The global decay parameters for U+2 DNA duplex in Mg^{2+} buffer at room temperature excited at 315 nm.

Global lifetimes/ ns: τ_1=0.04, τ_2=0.46, τ_3=2.5, τ_4=8.2. Global χ^2=1.083

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.88	0.06	0.04	0.02	1.075
380	0.86	0.08	0.04	0.02	1.067
395	0.84	0.09	0.05	0.02	1.106

Table A38 The global decay parameters for U+4 DNA duplex in Ca^{2+} buffer at room temperature excited at 315 nm.

Global lifetimes/ ns: τ_1=0.04, τ_2=0.44, τ_3=2.4, τ_4=8.2. Global χ^2=1.087

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.89	0.06	0.03	0.02	1.080
380	0.86	0.08	0.04	0.02	1.104
395	0.84	0.09	0.05	0.02	1.078

Table A39 The global decay parameters for U+4 DNA duplex in Mg^{2+} buffer at room temperature excited at 315 nm.

Global lifetimes/ ns: τ_1=0.05, τ_2=0.55, τ_3=2.4, τ_4=6.9. Global χ^2=1.053

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.70	0.15	0.11	0.04	1.073
380	0.68	0.16	0.11	0.05	1.039
395	0.67	0.17	0.11	0.05	1.048

Table A40 The global decay parameters for primer single strand in Ca^{2+} buffer at room temperature excited at 315 nm.

Global lifetimes/ns: τ_1=0.05, τ_2=0.54, τ_3=2.3, τ_4=6.7. Global χ^2=1.025

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.70	0.16	0.10	0.04	1.068
380	0.68	0.17	0.11	0.04	0.996
395	0.67	0.17	0.11	0.05	1.010

Table A41 The global decay parameters for primer single strand in Mg^{2+} buffer at room temperature excited at 315 nm.

Global lifetimes/ ns: τ_1=0.17, τ_2=0.94, τ_3=3.6, τ_4=9.0. Global χ^2=1.067

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.42	0.29	0.19	0.10	1.096
380	0.40	0.30	0.20	0.10	1.058
395	0.37	0.29	0.23	0.11	1.048

Table A42 The global decay parameters for U+2-exo$^-$ complex in Ca^{2+} buffer at room temperature excited at 315 nm.

Global lifetimes/ ns: τ_1=0.17, τ_2=0.99, τ_3=3.7, τ_4=8.7. Global χ^2=1.045

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.40	0.23	0.23	0.14	1.053
380	0.35	0.25	0.25	0.15	1.018
395	0.35	0.25	0.26	0.14	1.063

Table A43 The global decay parameters for U+2-exo$^-$ complex in Mg^{2+} buffer at room temperature excited at 315 nm.

Global lifetimes/ ns: τ_1=0.06, τ_2=0.67, τ_3=3.6, τ_4=9.5. Global χ^2=1.086

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.64	0.12	0.15	0.09	1.072
380	0.60	0.11	0.17	0.12	1.070
395	0.51	0.16	0.23	0.10	1.115

Table A44 The global decay parameters for U+4-exo$^-$ complex in Ca^{2+} buffer at room temperature excited at 315 nm.

Global lifetimes/ ns: τ_1=0.06, τ_2=0.67, τ_3=3.5, τ_4=9.4. Global χ^2=1.072

Emission wavelength/ nm	A_1	A_2	A_3	A_4	Local χ^2
365	0.66	0.11	0.15	0.08	1.094
380	0.60	0.12	0.17	0.11	1.080
395	0.55	0.15	0.21	0.10	1.047

Table A45 The global decay parameters for U+4-exo$^-$ complex in Mg^{2+} buffer at room temperature excited at 315 nm.

Appendix B

A Detailed Description of Archaeal Family-B DNA Polymerase and its Function

Russell *et al.*[1] investigated the interplay between the 3'-5' proofreading exonuclease activity and binding of the deaminated bases uracil (U) and hypoxanthine (H) with the family-B DNA polymerase from Pyrococcus furiosus (Pfu-Pol) by exonuclease and primer-template extension assays. Pfu-Pol was first identified by Uemori *et al.*[2] in 1997 and was found to have a strong 3'-5' exonucleolytic activity and a characteristic template-primer preference[2]. Russell *et al.*[1] found that the exonuclease and polymerase activities of the wild type Pfu-Pol (WT) were extremely slow when U was four bases ahead of the primer-template junction (+4 position). However, polymerization can still add two bases at extremely slow speed to give a strained conformation, in which the primer-template junction and the U binding site are very close. The exonuclease activity of the WT is strongly stimulated when U is at +3, +2, +1 and 0 positions ahead of the primer-template junction. They suggested that DNA with U at +3, +2, +1 and 0 positions ahead of the replication fork was bound to the DNA polymerase in an editing conformation with the 3' side of the primer strand unwinding to generate a conformational change, in which the single stranded region goes into the exonuclease active site. This change restores the spacing to four bases by unwinding the terminal bases in the primer, and therefore, relieves the strain. As a consequence, the distance between U and the primer-template junction is reinstated to the +4 position. The single stranded bases are removed by the 3'-5' exonuclease. The D215A point mutant of Pfu-Pol (exo^-), which lacks the exonuclease activity, was used to study the influence of the 3'-5' exonuclease activity on the ability of Pfu-Pol to replicate beyond the template strand deaminated bases, by extension assays. The result shows that exo^- is more proficient at extension than the WT. As a consequence, more full-length DNA with U is left in DNA-exo^- complex than in DNA-WT complex.

In order to elucidate how the archaeal family-B polymerase distinguishes U from the four DNA natural bases in the template single stranded region, and how the DNA interacts with the archaeal DNA polymerase, the X-ray crystal structure of Thermococcus gorgonarius (Tgo-Pol) DNA polymerase complexed with a DNA primer-template, containing U at +4 position (U+4) in the template single stranded region (sequence shown in Figure B1 (a)), was measured by Firbank *et al.*[3]. The crystal structure of the U+4-Tgo-Pol complex can mimic that of U+4-Pfu-Pol complex properly due to 80% of the amino acid sequence being the same between these two archaeal family-B DNA polymerases. The crystal structure of U+4-Tgo-Pol complex shows that the polymerase has five domains: exonuclease, thumb, palm, fingers and N terminal, which are shown in Figure B1 (b). Upon

binding to U+4, the thumb domain moves towards the palm domain, resulting in the distance between the thumb and exonuclease domains significantly increasing. The positions of the exonuclease, most of the palm, fingers and N terminal domains show nearly no change in the presence of DNA. The DNA double stranded region is bound to the thumb region of DNA polymerase. The binding pocket of U is located in the N-terminal domain. The N-terminal domain has little conformational change after binding U into the pocket, indicating that the pocked is formed before encountering any U during DNA replication.

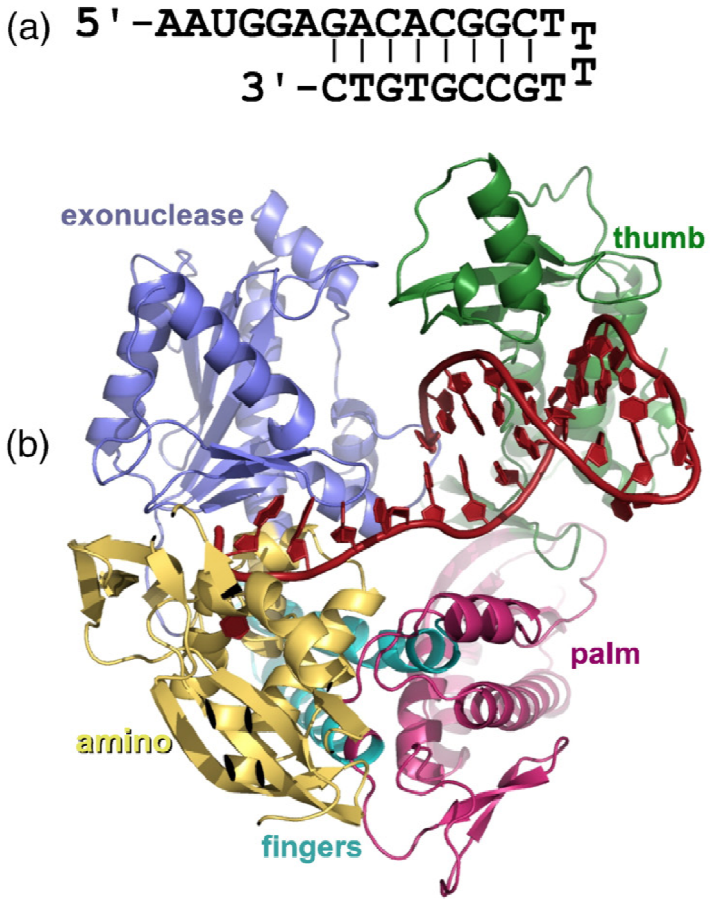
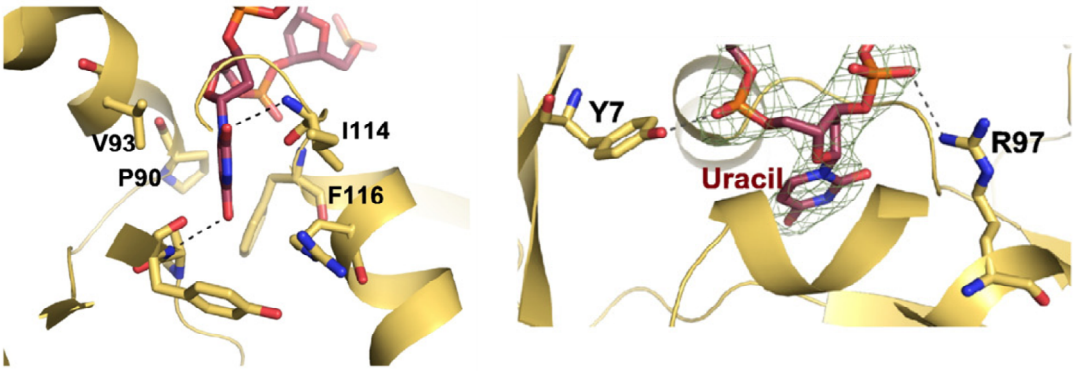


Figure B1 (a) The base sequence of primer-template (U+4) used to mimic DNA duplex and (b) the crystal structure of U+4-Tgo-Pol complex. Exonuclease domain is lilac, thumb domain is green, palm domain is magenta, finger domain is cyan, N-terminal domain is yellow and DNA is red. Reproduced from reference 3.

The U nucleotide is flipped to the binding pocked in *syn* conformation and stabilized by hydrogen bonds formed between the O_2 and O_4 atoms of U and the hydrogens from amide of Ile 114 and Tyr 37[3], respectively, as shown in Figure B2 (1). Val 93 forms π-stacking interaction with the flipped U, which is similar to the stacking interactions seen in DNA duplex. Pro 36, Pro 90 and Phe 116 are near to C_5 of U. The flanking phosphates of U are also stabilized by forming hydrogen bonds with Tyr 7 and Arg 97, respectively, as illustrated in Figure B2 (2). The backbone of DNA

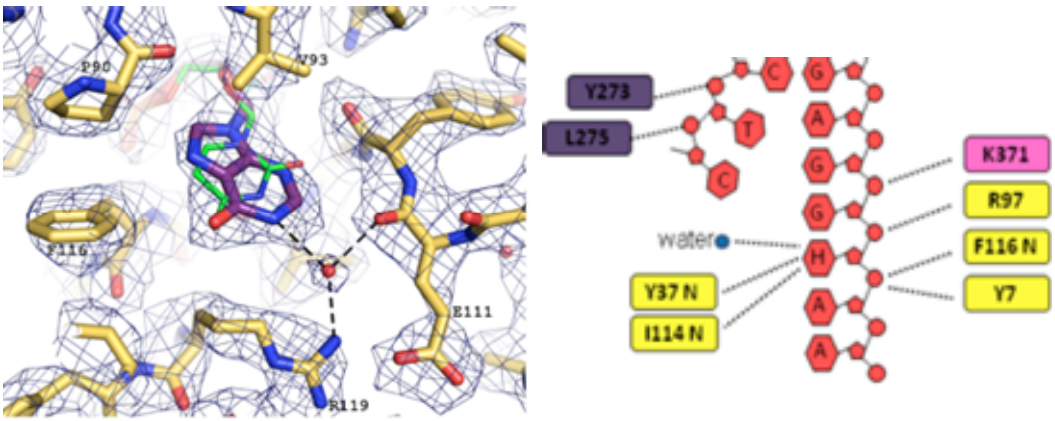
is remarkably distorted by the flipped U. The four DNA natural bases are excluded from the binding pocket by making use of hydrogen bonding and steric rejection.



(1) (2)

Figure B2 (1) The amino acids near the binding pocket of U in Pgo-Pol. Hydrogen bonding formed between two oxygen atoms of U and Ile 114 and Tyr 37, respectively. Val 93 forms π-stacking interaction with the flipped U, and Pro 36, Pro 90 and Phe 116 are near to C_5 of U. (2) Hydrogen bonds formed between 5' phosphate and Tyr 7, and 3' phosphate and Arg 97, respectively. Reproduced from reference 3.

Later, Killelea *et al.*[4] investigated the X-ray crystal structure of Tgo-Pol DNA polymerase in complex with a DNA primer-template containing H at +2 position (H+2) in the template single stranded region. The crystal structure of the H+2-Tgo-Pol complex shows that H is also in the thumb domain that is similar to U. However, the separation between thumb and exonucloease domains of Tgo-Pol is larger with H+2 than that with U+4. H is in *anti* conformation in the DNA polymerase binding pocket, which is identical to the one accommodating U. It is similar to U in the DNA polymerase that the binding of H barely changes the structure of the pocket and the amino acids involved for stabilizing the flipped H and DNA backbone. The amino acids, which hydrogen bond to the flipped H and its flanking phosphates, are the same as those bonding with U and flanking phosphates, as shown in Figure B3 (1). Additionally, the selectivity for this deaminated base is increased through a water-mediated hydrogen bond network linking of Arg 119 and Glu 111 of polymerase and N_1 of H. After DNA binding to the DNA polymerase and accompanying H at +2 position in the binding pocket, Arg 247 inserts between the fifth guanine (G) and the sixth adenine (A) from 5' side in the template and acts as a wedge, as shown in Figure B3 (2). In addition, Tyr 261 inserts into the second base-pair from replication fork to prevent the two bases of the 3' terminus of the primer re-annealing with the template. Therefore, the two bases of the 3' terminus of the primer become single stranded.



(1) (2)

Figure B3 (1) Superimposition of H (purple) and U (blue, from reference 3) in the binding pocket of Tgo-Pol. The water interacts through hydrogen bonding with O atom of E111, N_1 atom of H and side chain of R119. (2) The interactions of part of H+2 upon Pgo-Pol binding. H+2 is red, palm is pink, N terminal is yellow and exonuclease is blue. Reproduced from reference 4.

References

(1) Russell, H. J.; Richardson, T. T.; Emptage, K.; Connolly, B. A. *Nucleic Acids Res.* **2009**, *37*, 7603-7611.

(2) Uemori, T.; Sato, Y.; Kato, I.; Doi, H.; Ishino, Y. *Genes Cells* **1997**, *2*, 499-512.

(3) Firbank, S. J.; Wardle, J.; Heslop, P.; Lewis, R. J.; Connolly, B. A. *J. Mol. Biol.* **2008**, *381*, 529-539.

(4) Killelea, T.; Ghosh, S.; Tan, S. S.; Heslop, P.; Firbank, S. J.; Kool, E. T.; Connolly, B. A. *Biochemistry* **2010**, *49*, 5772-5781.

Appendix C

Structures and Results for Computational Calculation

Bond length	T/ Å				^{Se}T/ Å			
	HF	B3LYP	MP2/gas	MP2/solvated	HF	B3LYP	MP2/gas	MP2/solvated
C_4-O/Se	1.191	1.217	1.222	1.234	1.805	1.816	1.797	1.825
C_4-C_5	1.471	1.468	1.466	1.456	1.453	1.447	1.450	1.436
C_5-C	1.502	1.500	1.499	1.500	1.505	1.502	1.500	1.500
C_5-C_6	1.329	1.349	1.357	1.360	1.338	1.358	1.364	1.371
C_6-N_1	1.378	1.380	1.382	1.375	1.369	1.375	1.377	1.364
N_1-C_2	1.367	1.388	1.387	1.372	1.365	1.387	1.386	1.374
C_2-O	1.190	1.213	1.217	1.230	1.188	1.212	1.217	1.226
C_2-N_3	1.370	1.384	1.388	1.381	1.376	1.387	1.390	1.386
N_3-C_4	1.387	1.407	1.405	1.394	1.360	1.384	1.388	1.372

Table C1 The summary of all the bond lengths of ground-state equilibrium geometries of T and ^{Se}T in the gas phase optimised using B3LYP, HF, and MP2 with 6-311+G(d,p) basis set and in solution (water) calculated using SCRF-PCM/MP2/6-311+G(d,p).

Bond length	G/ Å				^{Se}G/ Å			
	HF	B3LYP	MP2/gas	MP2/solvated	HF	B3LYP	MP2/gas	MP2/solvated
C_6-O/Se	1.189	1.213	1.217	1.235	1.816	1.815	1.794	1.838
C_6-N_1	1.416	1.439	1.432	1.405	1.383	1.404	1.402	1.377
N_1-C_2	1.358	1.370	1.373	1.375	1.355	1.369	1.371	1.374
C_2-N	1.362	1.375	1.386	1.371	1.356	1.371	1.384	1.365
C_2-N_3	1.286	1.308	1.308	1.317	1.293	1.312	1.312	1.323
N_3-C_4	1.355	1.356	1.365	1.360	1.342	1.350	1.359	1.348
C_4-C_5	1.367	1.393	1.396	1.397	1.381	1.401	1.402	1.409
C_5-C_6	1.435	1.438	1.445	1.435	1.403	1.417	1.429	1.412
C_4-N_9	1.353	1.369	1.370	1.367	1.347	1.366	1.369	1.363
N_9-C_8	1.375	1.385	1.376	1.369	1.381	1.387	1.377	1.371
C_8-N_7	1.276	1.304	1.324	1.330	1.272	1.303	1.323	1.328
N_7-C_5	1.378	1.380	1.377	1.379	1.381	1.381	1.376	1.378

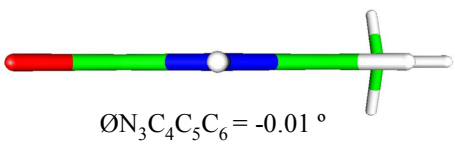
Table C2 The summary of all the bond lengths of ground-state equilibrium geometries of G and ^{Se}G in the gas phase optimised using B3LYP, HF, and MP2 with 6-311+G(d,p) basis set and in solution (water) calculated using SCRF-PCM/MP2/6-311+G(d,p).

Bond angle	T/ °				^{Se}T/ °			
	HF	B3LYP	MP2/gas	MP2/solvated	HF	B3LYP	MP2/gas	MP2/solvated
O/Se-C_4-C_5	124.6	125.1	124.9	124.8	124.9	125.1	124.9	124.5
N_3-C_4-C_5	115.0	114.6	114.3	114.9	115.7	115.5	114.9	115.8
O/Se-C_4-N_3	120.4	120.3	120.7	120.3	119.4	119.4	120.2	119.7
C_4-C_5-C	118.2	118.1	118.1	118.8	121.3	121.0	120.7	121.5
C-C_5-C_6	124.4	123.8	123.7	123.2	121.7	121.2	121.2	120.8
C_4-C_5-C_6	117.5	118.0	118.2	118.0	117.1	117.8	118.1	117.7
C_5-C_6-H	122.2	122.2	122.2	122.6	121.9	122.1	122.2	122.6
H-C_6-N_1	114.7	115.0	115.2	115.0	114.8	115.2	115.4	115.2
C_5-C_6-N_1	123.1	122.8	122.6	122.4	123.2	122.7	122.4	122.2
C_6-N_1-H	120.8	120.9	120.4	119.8	121.0	121.1	120.7	119.8
C_2-N_1-H	115.8	115.3	114.8	116.6	116.0	115.5	115.1	116.6
C_6-N_1-C_2	123.4	123.8	123.6	123.6	123.0	123.4	123.5	123.6
N_1-C_2-N_3	113.4	112.6	112.3	113.4	113.1	112.4	112.2	112.8
N_1-C_2-O	123.2	123.3	123.6	123.7	123.8	123.6	124.0	124.2
O-C_2-N_3	123.3	124.2	124.0	122.9	123.0	124.0	123.8	123.0
C_2-N_3-C_4	127.6	128.1	128.1	127.7	127.9	128.2	128.4	127.8
C_2-N_3-H	115.8	115.7	115.0	115.6	114.9	115.0	114.4	114.5
C_4-N_3-H	116.6	116.2	115.8	116.6	117.2	116.8	116.5	117.7

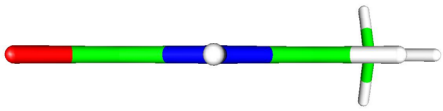
Table C3 The summary of all the bond angles of ground-state equilibrium geometries of T and ^{Se}T in the gas phase optimised using B3LYP, HF, and MP2 with 6-311+G(d,p) basis set and in solution (water) calculated using SCRF-PCM/MP2/6-311+G(d,p).

Bond angle	G/ °				^{Se}G/ °			
	HF	B3LYP	MP2/gas	MP2/solvated	HF	B3LYP	MP2/gas	MP2/solvated
O/Se-C_6-N_1	119.2	119.2	120.1	120.8	118.8	119.3	120.6	120.6
C_5-C_6-N_1	109.7	109.5	109.0	110.4	111.2	110.9	110.0	112.0
C_5-C_6-O/Se	131.1	131.1	130.8	128.8	130.0	129.9	129.4	127.3
C_6-N_1-H	114.2	113.5	113.7	115.3	114.6	113.9	114.2	116.4
H-N_1-C_2	119.5	119.9	118.6	118.3	119.3	119.8	118.7	117.6
C_6-N_1-C_2	126.2	126.5	126.8	126.4	126.0	126.2	126.8	126.0
N_1-C_2-N	116.3	117.0	115.8	116.1	116.6	117.0	115.8	116.5
N_3-C_2-N	119.9	119.6	120.0	120.1	120.6	119.6	120.0	120.2
N_1-C_2-N_3	123.8	123.4	124.1	123.7	123.5	123.4	124.1	123.2
C_4-N_3-C_2	112.8	112.8	111.7	112.3	112.8	112.7	111.6	112.7
C_5-C_4-N_3	128.9	129.1	129.5	128.5	128.0	128.3	128.7	127.8
N_9-C_4-N_3	125.7	125.9	125.7	126.5	126.4	126.4	126.2	127.2
N_9-C_4-C_5	105.4	105.0	104.8	104.9	105.6	105.3	105.0	105.1
N_7-C_5-C_6	130.9	130.4	129.8	130.1	131.6	131.1	130.1	127.3
C_6-C_5-C_4	118.6	118.7	118.6	118.6	118.4	118.6	118.7	118.3
N_7-C_5-C_4	110.5	110.8	111.5	111.2	109.9	110.3	111.2	110.8
C_5-N_7-C_8	104.8	104.8	103.9	103.8	105.0	105.0	104.0	103.8
N_7-C_8-N_9	112.8	112.6	112.9	112.8	113.1	112.9	113.2	113.3
H-C_8-N_9	121.5	121.7	121.8	121.8	121.2	121.5	121.7	121.5
H-C_8-N_7	125.7	125.7	125.3	125.5	125.7	125.6	125.1	125.2
C_8-N_9-C_4	106.4	106.7	106.9	107.2	106.3	106.5	106.6	106.0
C_4-N_9-H	126.0	125.6	125.6	126.4	126.1	125.8	125.8	126.6
C_8-N_9-H	127.6	127.7	127.5	126.4	127.6	127.7	127.6	126.5

Table C4 The summary of all the bond angles of ground-state equilibrium geometries of G and ^{Se}G in the gas phase optimised using B3LYP, HF, and MP2 with 6-311+G(d,p) basis set and in solution (water) calculated using SCRF-PCM/MP2/6-311+G(d,p).

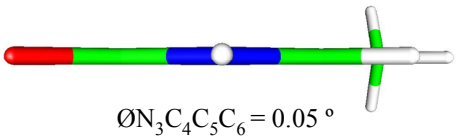


ØN$_3$C$_4$C$_5$C$_6$ = -0.01 °
ØC$_4$C$_5$C$_6$N$_1$ = 0.00 °
ØC$_5$C$_6$N$_1$C$_2$ = 0.01 °
ØC$_6$N$_1$C$_2$N$_3$ = -0.01 °
ØN$_1$C$_2$N$_3$C$_4$ = 0.00 °
ØC$_2$N$_3$C$_4$C$_5$ = 0.01 °
(1)

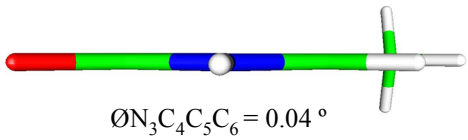


ØN$_3$C$_4$C$_5$C$_6$ = 0.04 °
ØC$_4$C$_5$C$_6$N$_1$ = -0.02 °
ØC$_5$C$_6$N$_1$C$_2$ = 0.00 °
ØC$_6$N$_1$C$_2$N$_3$ = -0.01 °
ØN$_1$C$_2$N$_3$C$_4$ = 0.04 °
ØC$_2$N$_3$C$_4$C$_5$ = -0.05 °
(2)

Figure C1 The side view and torsion angles of ground-state equilibrium geometries of (1) T and (2) ^{Se}T in the gas phase optimised using B3LYP/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen, and pink for selenium.

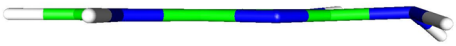


$ØN_3C_4C_5C_6 = 0.05$ °
$ØC_4C_5C_6N_1 = 0.01$ °
$ØC_5C_6N_1C_2 = -0.07$ °
$ØC_6N_1C_2N_3 = 0.06$ °
$ØN_1C_2N_3C_4 = 0.00$ °
$ØC_2N_3C_4C_5 = -0.06$ °
(1)

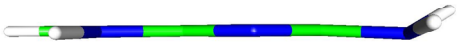


$ØN_3C_4C_5C_6 = 0.04$ °
$ØC_4C_5C_6N_1 = -0.00$ °
$ØC_5C_6N_1C_2 = -0.04$ °
$ØC_6N_1C_2N_3 = 0.03$ °
$ØN_1C_2N_3C_4 = 0.01$ °
$ØC_2N_3C_4C_5 = -0.05$ °
(2)

Figure C2 The side view and torsion angles of ground-state equilibrium geometries of (1) T and (2) ^{Se}T in the gas phase optimised using HF/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen, and pink for selenium.

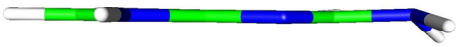


Six-member ring
$ØC_5C_6N_1C_2 = 0.09$ °
$ØC_6N_1C_2N_3 = 0.16$ °
$ØN_1C_2N_3C_4 = -0.78$ °
$ØC_2N_3C_4C_5 = 1.31$ °
$ØN_3C_4C_5C_6 = -1.12$ °
$ØC_4C_5C_6N_1 = 0.32$ °
Five-member ring
$ØC_8N_7C_5C_4 = -0.07$ °
$ØN_7C_5C_4N_9 = 0.10$ °
$ØC_5C_4N_9C_8 = -0.08$ °
$ØC_4N_9C_8N_7 = 0.04$ °
$ØN_9C_8N_7C_5 = 0.02$ °
(1)

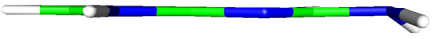


Six-member ring
$ØC_5C_6N_1C_2 = -0.75$ °
$ØC_6N_1C_2N_3 = 0.48$ °
$ØN_1C_2N_3C_4 = 0.46$ °
$ØC_2N_3C_4C_5 = -1.14$ °
$ØN_3C_4C_5C_6 = 0.86$ °
$ØC_4C_5C_6N_1 = 0.13$ °
Five-member ring
$ØC_8N_7C_5C_4 = 0.02$ °
$ØN_7C_5C_4N_9 = -0.04$ °
$ØC_5C_4N_9C_8 = 0.05$ °
$ØC_4N_9C_8N_7 = -0.04$ °
$ØN_9C_8N_7C_5 = 0.01$ °
(2)

Figure C3 The side view and torsion angles of ground-state equilibrium geometries of (1) G and (2) ^{Se}G in the gas phase optimised using B3LYP/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen, and pink for selenium.

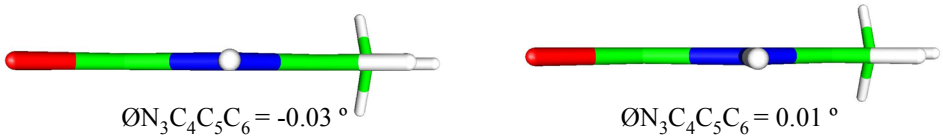


Six-member ring
$ØC_5C_6N_1C_2 = -0.30$ °
$ØC_6N_1C_2N_3 = 0.57$ °
$ØN_1C_2N_3C_4 = -0.83$ °
$ØC_2N_3C_4C_5 = 1.04$ °
$ØN_3C_4C_5C_6 = -0.87$ °
$ØC_4C_5C_6N_1 = 0.41$ °
Five-member ring
$ØC_8N_7C_5C_4 = -0.07$ °
$ØN_7C_5C_4N_9 = 0.10$ °
$ØC_5C_4N_9C_8 = -0.08$ °
$ØC_4N_9C_8N_7 = 0.04$ °
$ØN_9C_8N_7C_5 = 0.02$ °
(1)



Six-member ring
$ØC_5C_6N_1C_2 = 0.28$ °
$ØC_6N_1C_2N_3 = -0.11$ °
$ØN_1C_2N_3C_4 = -0.42$ °
$ØC_2N_3C_4C_5 = 0.84$ °
$ØN_3C_4C_5C_6 = -0.69$ °
$ØC_4C_5C_6N_1 = 0.08$ °
Five-member ring
$ØC_8N_7C_5C_4 = -0.04$ °
$ØN_7C_5C_4N_9 = 0.01$ °
$ØC_5C_4N_9C_8 = -0.06$ °
$ØC_4N_9C_8N_7 = 0.09$ °
$ØN_9C_8N_7C_5 = -0.08$ °
(2)

Figure C4 The side view and the torsion angles of ground-state equilibrium geometries of (1) G and (2) ^{Se}G in the gas phase optimised using HF/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen, and pink for selenium.



$ØN_3C_4C_5C_6 = -0.03$ º
$ØC_4C_5C_6N_1 = 0.01$ º
$ØC_5C_6N_1C_2 = 0.02$ º
$ØC_6N_1C_2N_3 = -0.02$ º
$ØN_1C_2N_3C_4 = -0.01$ º
$ØC_2N_3C_4C_5 = 0.03$ º
(1)

$ØN_3C_4C_5C_6 = 0.01$ º
$ØC_4C_5C_6N_1 = -0.01$ º
$ØC_5C_6N_1C_2 = 0.01$ º
$ØC_6N_1C_2N_3 = -0.01$ º
$ØN_1C_2N_3C_4 = 0.00$ º
$ØC_2N_3C_4C_5 = -0.00$ º
(2)

Figure C5 The side view and torsion angles of ground-state equilibrium geometries of (1) T and (2) ^{Se}T in solution (water) optimised using SCRF-PCM/MP2/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen, and pink for selenium.

Six-member ring
$ØC_5C_6N_1C_2 = 2.55$ º
$ØC_6N_1C_2N_3 = -1.40$ º
$ØN_1C_2N_3C_4 = -0.02$ º
$ØC_2N_3C_4C_5 = -0.10$ º
$ØN_3C_4C_5C_6 = 1.49$ º
$ØC_4C_5C_6N_1 = -2.43$ º
Five-member ring
$ØC_8N_7C_5C_4 = 0.45$ º
$ØN_7C_5C_4N_9 = -0.47$ º
$ØC_5C_4N_9C_8 = 0.29$ º
$ØC_4N_9C_8N_7 = -0.02$ º
$ØN_9C_8N_7C_5 = -0.26$ º
(1)

Six-member ring
$ØC_5C_6N_1C_2 = 2.39$ º
$ØC_6N_1C_2N_3 = -1.13$ º
$ØN_1C_2N_3C_4 = -0.04$ º
$ØC_2N_3C_4C_5 = -0.37$ º
$ØN_3C_4C_5C_6 = 1.82$ º
$ØC_4C_5C_6N_1 = -2.55$ º
Five-member ring
$ØC_8N_7C_5C_4 = 0.51$ º
$ØN_7C_5C_4N_9 = -0.56$ º
$ØC_5C_4N_9C_8 = 0.36$ º
$ØC_4N_9C_8N_7 = -0.06$ º
$ØN_9C_8N_7C_5 = -0.28$ º
(2)

Figure C6 The side view and torsion angles of ground-state equilibrium geometries of (1) G and (2) ^{Se}G in solution (water) optimised using SCRF-PCM/MP2/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen, and pink for selenium.

Bond length	Planar/ Å	Buckled/ Å	Bond angle	Planar/ º	Buckled/ º
C_4-O/Se	1.222	1.222	O/Se-C_4-C_5	125.0	124.9
C_4-C_5	1.466	1.466	N_3-C_4-C_5	114.3	114.3
C_5-C	1.499	1.499	O/Se-C_4-N_3	120.7	120.7
C_5-C_6	1.357	1.357	C_4-C_5-C	118.0	118.1
C_6-N_1	1.380	1.382	C-C_5-C_6	123.7	123.7
N_1-C_2	1.386	1.387	C_4-C_5-C_6	118.3	118.2
C_2-O	1.218	1.217	C_5-C_6-H	122.2	122.2
C_2-N_3	1.386	1.388	H-C_6-N_1	115.3	115.2
N_3-C_4	1.405	1.405	C_5-C_6-N_1	122.5	122.6
-	-	-	C_6-N_1-H	120.8	120.4
-	-	-	C_2-N_1-H	115.2	114.8
-	-	-	C_6-N_1-C_2	124.0	123.6
-	-	-	N_1-C_2-N_3	112.3	112.3
-	-	-	N_1-C_2-O	123.6	123.6
-	-	-	O-C_2-N_3	124.1	124.0
-	-	-	C_2-N_3-C_4	128.6	128.1
-	-	-	C_2-N_3-H	115.3	115.0
-	-	-	C_4-N_3-H	116.1	115.8

Table C5 The bond lengths and bond angles of the planar and buckled ground-state equilibrium geometry of T in the gas phase optimised using MP2/6-311+G(d,p).

T				G			
Bond length/ Å		Bond angle/ º		Bond length/ Å		Bond angle/ º	
C_4-O/Se	↑0.002	C_4-C_5-C_6	↑0.4	C_6-O	↓0.025	O-C_6-N_1	↓0.8
C_4-C_5	↑0.002	C_5-C_6-N_1	↓0.3	C_6-N_1	↑0.048	C_5-C_6-N_1	↓1.9
C_5-C	↑0.013	C_6-N_1-C_2	↑0.1	N_1-C_2	↓0.001	C_5-C_6-O	↑2.4
C_5-C_6	↑0.005	N_1-C_2-N_3	↓0.2	C_2-N	↑0.038	C_6-N_1-C_2	↑1.4
C_6-N_1	↑0.008	C_2-N_3-C_4	↑0.1	C_2-N_3	↓0.016	N_1-C_2-N	↑0.6
N_1-C_2	↑0.011	N_3-C_4-C_5	↓0.2	N_3-C_4	↑0.005	N_3-C_2-N	↓0.2
C_2-O	↑0.003	C-C_5-C_6	↓0.6	C_4-C_5	↑0.015	N_1-C_2-N_3	↓0.3
C_2-N_3	↑0.006	C_5-C_6-H	↓2.2	C_5-C_6	↑0.020	C_4-N_3-C_2	↑0.8
N_3-C_4	↑0.012	C_6-N_1-H	↓0.6	C_4-N_9	↓0.005	C_5-C_4-N_3	↑0.4
-	-	N_1-C_2-O	↑0.2	N_9-C_8	↑0.012	N_9-C_4-N_3	0
-	-	C_2-N_3-H	↑0.9	C_8-N_7	0	N_9-C_4-C_5	↓0.4
-	-	O-C_4-N_3	↓0.3	N_7-C_5	↓0.008	N_7-C_5-C_6	↓0.1
-	-	-	-	-	-	C_6-C_5-C_4	↓0.1
-	-	-	-	-	-	N_7-C_5-C_4	0
-	-	-	-	-	-	C_5-N_7-C_8	↑0.4
-	-	-	-	-	-	N_7-C_8-N_9	↓0.5
-	-	-	-	-	-	C_8-N_9-C_4	↑0.3
-	-	-	-	-	-	C_4-N_9-H	↓1.8
-	-	-	-	-	-	C_8-N_9-H	↑0.3

Table C6 The comparisons between the calculated bond lengths and bond angles of ground-state equilibrium geometry of T in the gas phase optimised using B3LYP/6-311+G(d,p) and the experimental structure obtained by the Joint Structural Analysis of the gas-phase electron diffraction (GED) and microwave (MW) data. The comparisons between the calculated bond lengths and bond angles of ground-state equilibrium geometry of G in the gas phase using B3LYP/6-311+G(d,p), and the estimated median bond lengths from a statistical survey of small molecules in the Cambridge Structural Database for which high-resolution X-ray and neutron crystal structures. ↑ means the calculated bond lengths and bond angles are bigger than the corresponding ones from experiment and ↓ represents the calculated bond lengths and bond angles are smaller than the corresponding ones from experiment.

Bond length	T/ Å		^{Se}T/ Å	
	CIS/gas	CIS/solvated	CIS/gas	CIS/solvated
C_4-O/Se	1.200	1.221	1.826	1.804
C_4-C_5	1.439	1.424	1.440	1.442
C_5-C	1.490	1.485	1.506	1.506
C_5-C_6	1.443	1.460	1.337	1.330
C_6-N_1	1.345	1.346	1.394	1.390
N_1-C_2	1.399	1.379	1.362	1.346
C_2-O	1.186	1.199	1.195	1.210
C_2-N_3	1.370	1.369	1.368	1.358
N_3-C_4	1.413	1.406	1.407	1.407

Table C7 The summary of all the bond lengths of the first excited-state equilibrium geometry of T and ^{Se}T in the gas phase optimised using CIS/6-311+G(d,p) and in solution (water) optimised using SCRF-PCM/CIS/6-311+G(d,p).

Bond length	G/ Å		^{Se}G/ Å	
	CIS/gas	CIS/solvated	CIS/gas	CIS/solvated
C_6-O/Se	1.177	1.214	1.854	1.850
C_6-N_1	1.406	1.416	1.421	1.407
N_1-C_2	1.348	1.374	1.363	1.344
C_2-N	1.313	1.322	1.372	1.328
C_2-N_3	1.338	1.330	1.281	1.322
N_3-C_4	1.294	1.316	1.366	1.330
C_4-C_5	1.424	1.454	1.362	1.407
C_5-C_6	1.460	1.420	1.432	1.418
C_4-N_9	1.358	1.389	1.358	1.349
N_9-C_8	1.341	1.335	1.366	1.361
C_8-N_7	1.322	1.316	1.284	1.293
N_7-C_5	1.317	1.347	1.375	1.373

Table C8 The summary of all the bond lengths of the first excited-state equilibrium geometry of G and ^{Se}G in the gas phase optimised using CIS/6-311+G(d,p) and in solution (water) optimised using SCRF-PCM/CIS/6-311+G(d,p).

Bond angle	T/ °		^{Se}T/ °	
	CIS/gas	CIS/solvated	CIS/gas	CIS/solvated
O/Se-C_4-C_5	127.0	126.8	123.9	125.4
N_3-C_4-C_5	113.3	114.3	116.4	116.7
O/Se-C_4-N_3	119.5	118.9	117.5	117.9
C_4-C_5-C	121.0	122.9	120.8	120.4
C-C_5-C_6	120.8	119.1	121.2	121.6
C_4-C_5-C_6	117.4	117.8	118.0	118.0
C_5-C_6-H	125.9	124.5	122.7	123.2
H-C_6-N_1	119.4	119.1	115.3	115.3
C_5-C_6-N_1	114.8	114.3	122.0	121.5
C_6-N_1-H	122.1	121.0	120.7	119.6
C_2-N_1-H	113.8	115.6	115.5	116.6
C_6-N_1-C_2	121.8	122.5	123.3	123.8
N_1-C_2-N_3	114.0	115.0	114.3	115.3
N_1-C_2-O	121.0	121.5	123.3	123.1
O-C_2-N_3	124.9	123.4	122.4	121.7
C_2-N_3-C_4	123.4	122.1	125.8	124.7
C_2-N_3-H	114.4	114.5	113.4	115.1
C_4-N_3-H	114.8	115.5	120.1	120.2

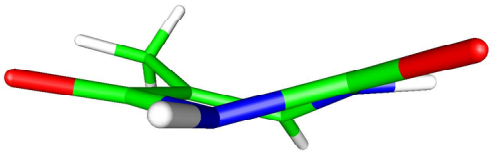
Table C9 The summary of all the bond angles of the first excited-state equilibrium geometry of T and ^{Se}T in the gas phase optimised using CIS/6-311+G(d,p) and in solution (water) optimised using SCRF-PCM/CIS/6-311+G(d,p).

Bond angle	G/ °		^{Se}G/ °	
	CIS/gas	CIS/solvated	CIS/gas	CIS/solvated
O/Se-C_6-N_1	121.4	120.0	118.6	118.1
C_5-C_6-N_1	110.2	110.8	110.5	111.7
C_5-C_6-O/Se	128.3	128.8	122.2	117.8
C_6-N_1-H	114.7	114.2	117.2	117.1
H-N_1-C_2	119.2	116.6	115.9	117.7
C_6-N_1-C_2	126.1	119.2	124.4	125.2
N_1-C_2-N	119.2	117.5	114.8	117.4
N_3-C_2-N	117.0	118.9	120.0	118.8
N_1-C_2-N_3	123.7	123.0	125.2	123.8
C_4-N_3-C_2	114.0	114.3	112.6	113.5
C_5-C_4-N_3	128.2	121.8	128.0	127.5
N_9-C_4-N_3	128.1	126.1	126.3	127.1
N_9-C_4-C_5	103.8	102.8	105.7	105.2
N_7-C_5-C_6	130.9	131.2	130.6	132.4
C_6-C_5-C_4	117.9	118.7	119.2	118.1
N_7-C_5-C_4	111.2	110.0	110.2	109.3
C_5-N_7-C_8	104.7	105.0	105.1	105.2
N_7-C_8-N_9	113.4	114.6	112.6	113.1
H-C_8-N_9	122.5	121.8	121.8	121.4
H-C_8-N_7	124.0	123.6	125.6	125.5
C_8-N_9-C_4	106.9	107.0	106.4	107.1
C_4-N_9-H	125.9	126.8	126.2	126.5
C_8-N_9-H	127.2	125.4	127.4	126.4

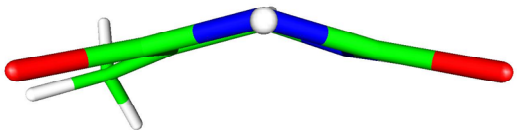
Table C10 The summary of all the bond angles of the first excited-state equilibrium geometry of G and ^{Se}G in the gas phase optimised using CIS/6-311+G(d,p) and in solution (water) using SCRF-PCM/CIS/6-311+G(d,p).

Bond length	Gas/ Å	Solvation/ Å	Bond angle	Gas/ °	Solvation/ °
C_6-O/Se	↓0.011	↓0.002	O/Se-C_6-N_1	-	-
C_6-N_1	↓0.026	↓0.022	C_5-C_6-N_1	-	-
N_1-C_2	↓0.051	↓0.014	C_5-C_6-O/Se	↓0.1	↓1
C_2-N	↓0.060	↓0.016	C_6-N_1-H	-	-
C_2-N_3	↓0.057	↑0.005	H-N_1-C_2	-	-
N_3-C_4	↑0.006	↓0.008	C_6-N_1-C_2	-	-
C_4-C_5	↓0.008	↓0.012	N_1-C_2-N	-	-
C_5-C_6	↑0.003	↑0.002	N_3-C_2-N	-	-
C_4-N_9	↓0.019	↓0.016	N_1-C_2-N_3	↑5.3	↑0.2
N_9-C_8	↓0.020	↓0.003	C_4-N_3-C_2	↑5.1	↓0.2
C_8-N_7	↑0.026	↑0.005	C_5-C_4-N_3	↑3.2	↑1.4
N_7-C_5	↓0.040	↓0.008	N_9-C_4-N_3	-	-
-	-	-	N_9-C_4-C_5	↓0.2	↑0.4
-	-	-	N_7-C_5-C_6	-	-
-	-	-	C_6-C_5-C_4	↓1.4	↑0.1
-	-	-	N_7-C_5-C_4	↑1.1	↓0.2
-	-	-	C_5-N_7-C_8	↓0.4	↓0.1
-	-	-	N_7-C_8-N_9	↓1.1	↓0.6
-	-	-	H-C_8-N_9	-	-
-	-	-	H-C_8-N_7	-	-
-	-	-	C_8-N_9-C_4	-	-
-	-	-	C_4-N_9-H	-	-
-	-	-	C_8-N_9-H	-	-

Table C11 The comparisons between the calculated bond lengths and bond angles of the first excited-state equilibrium geometry of G in the gas phase optimised using CIS/6-311+G(d,p) in this work and those at CIS/cc-pVDZ in the literature. The comparisons between the calculated bond lengths and bond angles of the first excited-state equilibrium geometry of G in solution (water) optimised using SCRF-PCM/CIS/6-311+G(d,p) in this work and those at IEF/CIS/cc-pVDZ. ↑ means the calculated bond length and bond angles in this work is bigger than the corresponding ones from literature and ↓ represents the calculated bond lengths and bond angles from this work is smaller than the corresponding ones from literature.

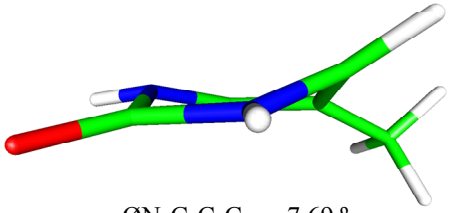


$ØN_3C_4C_5C_6 = 7.68$ º
$ØC_4C_5C_6N_1 = -36.40$ º
$ØC_5C_6N_1C_2 = 35.86$ º
$ØC_6N_1C_2N_3 = -5.29$ º
$ØN_1C_2N_3C_4 = -27.19$ º
$ØC_2N_3C_4C_5 = 25.02$ º
(1)

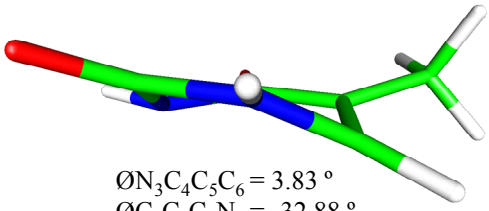


$ØN_3C_4C_5C_6 = -2.97$ º
$ØC_4C_5C_6N_1 = 31.47$ º
$ØC_5C_6N_1C_2 = -33.31$ º
$ØC_6N_1C_2N_3 = 5.04$ º
$ØN_1C_2N_3C_4 = 27.50$ º
$ØC_2N_3C_4C_5 = -27.46$ º
(2)

Figure C7 The side view and torsion angles of the first excited-state equilibrium geometry of T (1) in the gas phase (T_{gas}) optimised using CIS/6-311+G(d,p) and (2) in solution (water) (T_{sol}) using SCRF-PCM/CIS/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, and red for oxygen.



$ØN_3C_4C_5C_6 = -7.69$ º
$ØC_4C_5C_6N_1 = 36.39$ º
$ØC_5C_6N_1C_2 = -33.31$ º
$ØC_6N_1C_2N_3 = 5.28$ º
$ØN_1C_2N_3C_4 = 27.18$ º
$ØC_2N_3C_4C_5 = -25.01$ º
(1)

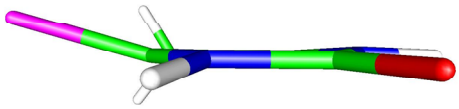


$ØN_3C_4C_5C_6 = 3.83$ º
$ØC_4C_5C_6N_1 = -32.88$ º
$ØC_5C_6N_1C_2 = -34.31$ º
$ØC_6N_1C_2N_3 = -5.20$ º
$ØN_1C_2N_3C_4 = -27.84$ º
$ØC_2N_3C_4C_5 = 27.42$ º
(2)

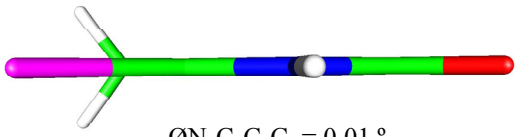
Figure C8 The side view of the first excited-state equilibrium geometry of T (1) in the gas phase optimised using CIS/HF/6-311+G(d,p) level, based on the optimised first excited-state equilibrium geometry of T in solution (water) (T_{sol}); and (2) in solution (water) optimised using SCRF-PCM/CIS/6-311+G(d,p), based on the optimised buckled first excited-state equilibrium geometry of T in the gas phase (T_{gas}). Green stands for carbon, blue for nitrogen, white for hydrogen, and red for oxygen.

Bond length/ Å	T_{gas} (c)	T_{gas}	T_{sol}(c)	T_{sol}	Bond angle/ °	T_{gas} (c)	T_{gas}	T_{sol}(c)	T_{sol}
C_4-O/Se	1.200	1.200	1.220	1.221	O/Se-C_4-C_5	127.0	127.0	126.9	126.8
C_4-C_5	1.439	1.439	1.425	1.424	N_3-C_4-C_5	113.3	113.3	114.1	114.3
C_5-C	1.490	1.490	1.486	1.485	O/Se-C_4-N_3	119.5	119.5	119.0	118.9
C_5-C_6	1.443	1.443	1.458	1.460	C_4-C_5-C	121.0	121.0	122.6	122.9
C_6-N_1	1.345	1.345	1.346	1.346	C-C_5-C_6	120.8	120.8	119.4	119.1
N_1-C_2	1.399	1.399	1.380	1.379	C_4-C_5-C_6	117.4	117.4	117.6	117.8
C_2-O	1.186	1.186	1.198	1.199	C_5-C_6-H	125.9	125.9	125.0	124.5
C_2-N_3	1.307	1.370	1.370	1.369	H-C_6-N_1	119.4	119.4	119.4	119.1
N_3-C_4	1.413	1.413	1.406	1.406	C_5-C_6-N_1	114.7	114.8	114.3	114.3
-	-	-	-	-	C_6-N_1-H	122.1	122.1	121.1	121.0
-	-	-	-	-	C_2-N_1-H	113.8	113.8	115.6	115.6
-	-	-	-	-	C_6-N_1-C_2	121.8	121.8	122.1	122.5
-	-	-	-	-	N_1-C_2-N_3	114.0	114.0	115.0	115.0
-	-	-	-	-	N_1-C_2-O	121.0	121.0	121.6	121.5
-	-	-	-	-	O-C_2-N_3	124.9	124.9	123.4	123.4
-	-	-	-	-	C_2-N_3-C_4	123.4	123.4	122.1	122.1
-	-	-	-	-	C_2-N_3-H	114.4	114.4	114.4	114.5
-	-	-	-	-	C_4-N_3-H	114.8	114.8	115.4	115.5

Table C12 The bond lengths and bond angles of two conformers of the first excited-state equilibrium geometry of T in the gas phase optimised using CIS/6-311+G(d,p) and in solution (water) using SCRF-PCM/CIS/6-311+G(d,p). (c) means the corresponding conformer.

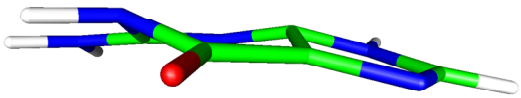


$ØN_3C_4C_5C_6$ = -1.38 °
$ØC_4C_5C_6N_1$ = -2.10 °
$ØC_5C_6N_1C_2$ = 6.05 °
$ØC_6N_1C_2N_3$ = -5.66 °
$ØN_1C_2N_3C_4$ = 1.96 °
$ØC_2N_3C_4C_5$ = 1.46 °
(1)



$ØN_3C_4C_5C_6$ = 0.01 °
$ØC_4C_5C_6N_1$ = 0.01 °
$ØC_5C_6N_1C_2$ = -0.00 °
$ØC_6N_1C_2N_3$ = -0.02 °
$ØN_1C_2N_3C_4$ = 0.04 °
$ØC_2N_3C_4C_5$ = -0.04 °
(2)

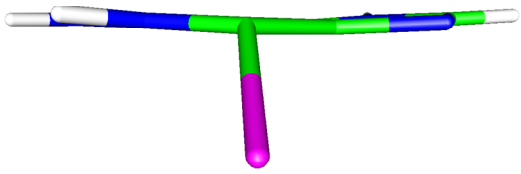
Figure C9 The side view and torsion angles of the first excited-state equilibrium geometry of ^{Se}T (1) in the gas phase optimised using CIS/6-311+G(d,p) and (2) in solution (water) using SCRF-PCM/CIS/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen and pink for selenium.



Six-member ring
$ØC_5C_6N_1C_2$ = 0.01 °
$ØC_6N_1C_2N_3$ = -0.00 °
$ØN_1C_2N_3C_4$ = 0.01 °
$ØC_2N_3C_4C_5$ = -0.02 °
$ØN_3C_4C_5C_6$ = 0.03 °
$ØC_4C_5C_6N_1$ = -0.02 °
Five-member ring
$ØC_8N_7C_5C_4$ = 0.04 °
$ØN_7C_5C_4N_9$ = -0.02 °
$ØC_5C_4N_9C_8$ = -0.01 °
$ØC_4N_9C_8N_7$ = 0.04 °
$ØN_9C_8N_7C_5$ = -0.05 °
(1)

Six-member ring
$ØC_5C_6N_1C_2$ = 29.45 °
$ØC_6N_1C_2N_3$ = -34.51 °
$ØN_1C_2N_3C_4$ = 6.09 °
$ØC_2N_3C_4C_5$ = -23.04 °
$ØN_3C_4C_5C_6$ = -25.26 °
$ØC_4C_5C_6N_1$ = -2.25 °
Five-member ring
$ØC_8N_7C_5C_4$ = -4.51 °
$ØN_7C_5C_4N_9$ = 7.41 °
$ØC_5C_4N_9C_8$ = -7.26 °
$ØC_4N_9C_8N_7$ = 5.24 °
$ØN_9C_8N_7C_5$ = -0.37 °
(2)

Figure C10 The torsion angles of the first excited-state equilibrium geometry of G (1) in the gas phase optimised using CIS/6-311+G(d,p) and (2) the side view and torsion angles of that in solution (water) optimised using SCRF-PCM/CIS/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, and red for oxygen.



Six-member ring
$ØC_5C_6N_1C_2$ = 1.82 °
$ØC_6N_1C_2N_3$ = 0.04 °
$ØN_1C_2N_3C_4$ = -2.67 °
$ØC_2N_3C_4C_5$ = 3.74 °
$ØN_3C_4C_5C_6$ = -1.96 °
$ØC_4C_5C_6N_1$ = -0.89 °
Five-member ring
$ØC_8N_7C_5C_4$ = 0.39 °
$ØN_7C_5C_4N_9$ = -0.60 °
$ØC_5C_4N_9C_8$ = 0.56 °
$ØC_4N_9C_8N_7$ = -0.35 °
$ØN_9C_8N_7C_5$ = -0.02 °
(1)

Six-member ring
$ØC_5C_6N_1C_2$ = 2.43 °
$ØC_6N_1C_2N_3$ = -1.77 °
$ØN_1C_2N_3C_4$ = -0.09 °
$ØC_2N_3C_4C_5$ = 1.05 °
$ØN_3C_4C_5C_6$ = -0.20 °
$ØC_4C_5C_6N_1$ = -1.43 °
Five-member ring
$ØC_8N_7C_5C_4$ = 0.67 °
$ØN_7C_5C_4N_9$ = -1.00 °
$ØC_5C_4N_9C_8$ = 0.92 °
$ØC_4N_9C_8N_7$ = -0.56 °
$ØN_9C_8N_7C_5$ = -0.08 °
(2)

Figure C11 The torsion angles of the first excited-state equilibrium geometry of ^{Se}G (1) in the gas phase optimised using CIS/6-311+G(d,p) and (2) the side view and torsion angles of that in solution (water) optimised using SCRF-PCM/CIS/6-311+G(d,p). Green stands for carbon, blue for nitrogen, white for hydrogen, red for oxygen and pink for selenium.

Based on the planar starting geometry			Based on the buckled starting geometry		
Excited-state	**Wavelength (scaled)/ nm**	**Oscillator strength**	**Excited-state**	**Wavelength (scaled)/ nm**	**Oscillator strength**
1	277	0.438	1	277	0.430
2	274	0	2	275	0.003
3	261	0.005	3	260	0.005
4	231	0.007	4	230	0.008
5	223	0	5	224	0.003
6	223	0	6	223	0
7	219	0.007	7	218	0.009
8	211	0.210	8	211	0.204
9	201	0.414	9	201	0.390
10	197	0.019	10	197	0.028
11	196	0	11	196	0.001
12	195	0.022	12	194	0.022
13	188	0.001	13	188	0.001
14	188	0.004	14	187	0.010
15	186	0.056	15	186	0.056

Table C13 The first fifteen absorption transition wavelengths (scaled by a factor of 0.72), along with their oscillator strengths, for T in the gas phase calculated using CIS/6-311+G(d,p) based on the planar starting geometry and buckled starting geometry optimised using MP2/6-311+G(d,p).

Conformer			T_{sol}		
Excited-state	**Wavelength (scaled)/ nm**	**Oscillator strength**	**Excited-state**	**Wavelength (scaled)/ nm**	**Oscillator strength**
1	408	0.276	1	398	0.292
2	277	0.015	2	276	0.017
3	260	0.006	3	259	0.004

Table C14 The first three fluorescence transition wavelengths (scaled by a factor of 0.72), along with their oscillator strengths, of T in solution (water) calculated using SCRF-PCM/CIS/6-311+G(d,p) based on the first excited-state equilibrium geometries of T in solution (water) (T_{sol}) and its conformer optimised using the same level of theory.

Appendix D

Lectures and Conferences

Postgraduate Lectures, Courses and Workshops

Year 1

Spectroscopy
Structures of biological macromolecules
Biophotonics
Biosensors
Statistics and data handling
Microscopy and imaging
General safety lecture
Fire safety awareness and extinguishers
Laser safety course
Enhancing English language and skills through public engagement
PhD academic paper writing
How to be an effective researcher
Time management
Informatics distinguished lecture
Producing a thesis using word
Word introduction: creating and formatting documents
Word intermediate: document layout and customization
Excel introduction: creating basic spreadsheets and charts
Excel advanced: manipulating and managing data
Managing bibliographies with Endnote
Basic image manipulation techniques using Photoshop
Preparing a presentation with PowerPoint
School of chemistry colloquia

Year 2

Introduction to biosafety
Safety requirements for genetic modification work
Microbiological safety cabinets
Transport of biological materials
Information resources and searching research literature for chemistry
Literature searching - chemistry
Proof reading
Effective writing - the writing process
Effective writing - grammar

Mapping your mind and reading for speed
School of chemistry colloquia

Year 3

Introduction to computational chemistry
Introduction to high-performance computing and Linux
Visualization software
Electronic structure software
Spectroscopic properties
Classical simulation software
Periodic electronic structure software
Crystallography workshop
Risk Assessment Lecture
Introduction to research funding
How to write a winning CV
Successful job hunting
Interviews: how to succeed
Preparing and printing large format posters
School of chemistry colloquia

Conferences, Meetings and Presentation Given

School of Chemistry, Physical Chemistry Section Meetings
Oral Presentations
Edinburgh, UK, February 2008, February 2009, March 2010

6th International Optical Society of America Network of Students
Poster Presentation
Glasgow, UK, July 2009

42nd IUPAC Congress
Oral Presentation
Glasgow, UK, July 2009

e-Science Data, Information and Knowledge Transformation 2010 Symposium
Poster Presentation
Edinburgh, UK, April 2010

ScotCHEM Computational Chemistry Symposium 2010
Poster Presentation
Glasgow, UK, April 2010

XXIII IUPAC Symposium on Photochemistry
Poster Presentation
Ferrara, Italy, July 2010